University of Alberta

Methodological Improvements in Protein Crystallography

by

Kirsty Velda Dunlop

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

in

Macromolecular Crystallography

Department of Medical Microbiology and Immunology

Edmonton, Alberta

Fall 2005

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

0-494-08633-5

Your file *Votre référence*
ISBN
Our file *Notre reference*
ISBN

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Abstract

Methodological improvements in protein x-ray crystallography are essential in this new age of structural proteomics. This paper-based thesis focuses on the study and development of methodological improvements in two important areas of crystallography, crystallization and crystal cryocooling. The first paper introduces the "dilution method", a new method of reducing protein consumption using a modified vapour diffusion protocol. Reducing protein consumption is of utmost importance to crystallographers because of the time, effort, and money that goes into producing pure protein. In this protocol, the protein and mother liquor in the crystallization drop are both diluted, while the mother liquor in the well remains undiluted. In addition, the dilution method circumvents some of the current problems associated with robotic crystallization screening trials including evaporation of the drop, high viscosity, and dispensing accuracy. Finally, the dilution method was found to be a promising crystal optimization method. The second paper introduces the "shared reservoir solution method", a vapor diffusion crystallization method where many different experiments share the same reservoir solution. The method significantly simplifies manual and robotic experiment setup, and reduces cost. Also, recent reports and our own results indicate that this method may actually increase crystallization success. Most importantly, the method allows for the design of a completely new open well crystallization plate. These plates are optically superior, and show significant advantages during automatic imaging and crystal detection and also allow for higher drop density. The third paper disects, for the first time, the

effects of cryocooling, resolution, and radiation damage on crystal structures. Our results show that although the high resolution cryogenic structures are more precise and more detailed, they also show systematic deviations from the room temperature structures especially at the protein surface. These differences can affect biological interpretations, especially because many important biological processes take place at the protein surface. In addition, the high intensity synchrotron beam often causes significant specific and non-specific radiation damage to the crystal structure. To overcome this, data collection strategies should be designed with this potential problem in mind.

Acknowledgements

Thank you to everyone involved in my studies, from supervisor, co-workers and fellow students to friends and family. You have enriched my life and helped me to achieve my goals.

I would like to thank my supervisor Dr. Bart Hazes for his invaluable guidance. His friendliness, positive outlook, and knowledge make him a truly amazing supervisor.

Thank you to Gerald Audette for helpful discussions regards the dilution method.
Thank you to Luke Price for programming the Honeybee robot to accept our SRS plates.
Thank you to Angela Brigley for maintaining a well ordered lab.
Thank you to Koto, Luke, John-Paul, Kevin, and Lisa who were always friendly faces in the lab.

Thank you to my husband Patrick Turc who is always there with fun, sparkling eyes, and loving songs of encouragement.

Lastly, I would like to thank my Mum for her unconditional love.

Table of Contents

Chapter 1: Introduction **1**

1.1 Structural Biology 1

1.2 Outline of X-Ray Crystallography Process 2

1.2.1 Protein Production 3

1.2.2 Protein Crystallization 4

1.2.3 Collecting Diffraction Data 4

1.2.4 Structure Solution 5

1.3 Crystallization 8

1.3.1 Supersaturation, Nucleation, Growth, and Growth Cessation 10

1.3.1.1 Supersaturation *10*

1.3.1.2 Nuclei Formation *11*

1.3.1.3 Crystal Growth *13*

1.3.1.4 Growth Cessation *15*

1.3.2 Crystallization Methods 16

1.3.2.1 Vapor Diffusion Technique *16*

1.3.2.2 Microbatch Crystallization Method *17*

1.3.2.3 Free Interface Diffusion and Microfluidics *20*

1.3.3 Screening 22

1.3.4 Optimization 28

1.4 Methods Improvements in Crystallography 29

1.4.1 Methods Improvements in Crystallization 30

1.4.1.1 Automation of the Crystallization Process *30*

1.4.1.2 Crystallization Trays *31*

1.4.1.3 Crystallization Robots *31*

1.4.1.4 Automatic Imaging 33
1.4.1.5 Automatic Classification 34
1.4.2 Methods Improvements in Data Collection 36
1.4.2.1 Increase in Brilliance of Synchrotrons 36
1.4.2.2 Ultra-high Resolution Structures 37
1.4.2.3 Cryocrystallography and Radiation Damage 40
1.4.2.3.1 History of Cryocrystallography 40
1.4.2.3.2 Theory of Radiation Damage 42
1.4.2.3.3 Minimizing Radiation Damage 47
1.4.2.3.4 Structural Changes Resulting From Flash Cooling 48
1.5 Conclusion to Introduction 49
1.6 References 49
Chapter 2: When Less Is More: A More Efficient Vapour Diffusion Protocol **59**
2.1 Introduction 60
2.1.1 Reducing Protein Consumption 60
2.1.2 The Dilution Method 61
2.2 Experimental 62
2.2.1 Protein Preparation 62
2.2.2 Crystal Growth 63
2.3 Results/Discussion 63
2.3.1 Effects of the Vapour Diffusion Method on Crystallization 63
2.3.2 Dilution and Traditional Vapour Diffusion Method Comparison 66
2.3.3 Advantages and Concerns of the Dilution Method 66
2.4 Conclusions 67
2.5 References 68

Chapter 3: A Modified Vapor Diffusion Crystallization Protocol That Uses A Common Dehydrating Agent..... 69

3.1 Introduction.....70

3.2 Experimental.....73

3.2.1 Protein Preparation.....73

3.2.2 Crystallization.....73

3.3 Results.....74

3.3.1 Choice of Precipitant and Concentration to Use as Shared Reservoir Solution.....74

3.3.2. Effects of Non-Aqueous Volatile Components in the Crystallization Drop.....76

3.3.3 Crystal Screening.....76

3.3.4 SRS Crystallization Plate.....78

3.3.5 Optical Quality of SRS Plates.....80

3.4 Discussion.....82

3.4.1 Choice of SRS Precipitant.....82

3.4.2 Effect of Volatile Reagents in the Crystallization Drop.....83

3.4.3 Choice of Reservoir Solution Dehydrating Strength.....83

3.4.4 Practical Advantages of SRS Plates.....85

3.5 Conclusions.....86

3.6 References.....86

Chapter 4: Pros and Cons of Cryocrystallography- Should We Also Collect a Room Temperature Data Set?..... 89

4.1 Introduction.....90

4.2 Materials and Methods.....92

4.2.1 Crystallization, Data Collection, and Data Processing.....92

4.2.2 Refinement 93
4.2.3 Hydrogen Atom Density 94
4.2.4 Hydration Shells 94
4.2.5 Program availability 94
4.3 Results and Discussion 95
4.3.1 Ultra-High Resolution Structure and Radiation Damage 95
4.3.2 Radiation Damage to the Disulphide Bond 95
4.3.3 Hydrogen Atom Density 97
4.3.4 Temperature-Induced Structural Changes 98
4.3.5 Structural Variation in the Main Chain 99
4.3.6 Structural Variation in the Side Chains 101
4.3.7 Differences in Hydration Structure 104
4.4 Conclusions 109
4.5 References 110
Chapter 5: Discussion 113
5.1 Discussion 113
5.2 When Less Is More: A More Efficient Vapor Diffusion Protocol 113
5.3 A Modified Vapor Diffusion Protocol That Uses A Universal Reservoir Solution 115
5.4 Pros and Cons of Cryocrystallography: Should We Also Collect a Room-Temperature Data Set? 119
5.5 Conclusion 123
5.6 References 124

List of Tables

Table 1.1 Parameters that effect crystallization (McPherson, 2004b). 23
Table 2.1 The effect of the dilution method on lysozyme crystal growth.....64
Table 2.2 The effect of the dilution method on glucose isomerase crystal growth. 64
Table 3.1 Evaluation of NaCl and PEG 3350 as Shared Reservoir Solutions.....75
Table 3.2 Comparison of the SRS and traditional methods in crystal screens. 77
Table 3.3 Comparison of the SRS and traditional plates using crystal screens. 80
Table 4.1 Batch statistics for the ultra-high 0.78 Å resolution data set.....93
Table 4.2 Diffraction data and refinement statistics.....99
Table 4.3 Structural deviations between pairs of PAK structures 101
Table 4.4 Multiple side-chain conformations of the six PAK structures...102
Table 4.5 Incomplete side chains for the six PAK structures.....104
Table 4.6 Number of water molecules and water shell distribution 105
Table 4.7 Structurally equivalent water molecules 107

Table of Figures

Figure 1.1 Steps in Structure Determination Using X-ray Crystallography...2
Figure 1. 2 Cryocooled crystal mounted in a loop....4
Figure 1. 3 Diffraction image....5
Figure 1. 4 Model building....7
Figure 1. 5 Bottlenecks in structure determination....8
Figure 1. 6 Phase diagram for protein solubility in solution as a function of precipitant and protein concentration....11
Figure 1. 7 In order for a protein to crystallize it must overcome an energy barrier....12
Figure 1. 8 Intermolecular contact areas and nucleus stability....13
Figure 1. 9 Tangential and crystal face growth....14
Figure 1. 10 Atomic force microscopy examples of crystal face growth mechanisms....15
Figure 1. 11 A vapor diffusion crystallization experiment....16
Figure 1. 12 Phase diagram for crystal growth in a vapor diffusion experiment....17
Figure 1. 13 Microbatch method pipetting sequence....18
Figure 1. 14 Phase diagram for microbatch crystallization experiments....19
Figure 1. 15 FID crystallization....20
Figure 1. 16 Phase diagram for free interface diffusion....21
Figure 1. 17 A microfluidic chip....21
Figure 1. 18 Screening Methods....25
Figure 1. 19 The success rate of crystallization....26
Figure 1. 20 Static deformation density of a peptide plane in crambin....37
Figure 1. 21 Hydrogen atom density....38

Figure 1. 22 Alternate conformations 39
Figure 1. 23 Carbon, nitrogen, and oxygen atoms can be differentiated at ultra-high resolution. 39
Figure 1. 24 Survey of cryo versus room temperature data collection. 41
Figure 1.25 Crystal cryocooling.....41
Figure 1.26 Categories of radiation damage to a crystal.....43
Figure 1.27 Effects of radiation damage on a crystal.....43
Figure 1.28 Increase in overall B-factors.....44
Figure 1.29 Unit-cell volume increase as a function of deposited radiation dose.....45
Figure 1.30 Difference Fourier maps showing time course cleavage of a disulphide bond 46
Figure 2. 1 Crystallization space diagram showing the path of mother liquor and protein concentrations during a vapour diffusion crystallization experiment.....62
Figure 2.2 Crystals of lysozyme and glucose isomerase. 65
Figure 3.1 SRS crystallization plate.....79
Figure 3.2 Comparison of crystals in SRS and commercial plates.....81
Figure 4.1 2Fo –Fc electron density for cysteine residues 129 and 142.....96
Figure 4.2 Hydrogen atom density.....98
Figure 4.3 Temperature related alternate side chain conformations..... 103
Figure 4.4 Superimposed cryocooled and room-temperature water molecules.....108
Figure 5.1 The dilution method at work..... ...115
Figure 5.2 SRS plate design.....117

Figure 5.3 SRS plate design..... 119

List of Abbreviations

EM	Electron Microscopy
FID	Free Interface Diffusion
GST	Glutathione S-Transferase
MAD	Multiwavelength Anomalous Dispersion
MIR	Multiple Isomorphous Replacement
MIRAS	Multiple Isomorphous Replacement with Anomalous Scattering
MPD	2-Methyl-2,4-pentanediol
NMR	Nuclear Magnetic Resonance
PEG	Polyethylene Glycol
PDB	Protein Data Bank
RIP	Radiation damage-Induced Phasing
SAD	Single wavelength Anomalous Dispersion
SIR	Single Isomorphous Replacement
SBS	Society of Biomolecular Screening
SRS	Shared Reservoir Solution

Chapter 1: Introduction

1.1 Structural Biology

Proteins are one of the most important players in the life of the cell. Their functions are numerous, including, catalyzing chemical reactions, controlling gene expression, acting as receptors of intercellular and intracellular signals, regulating the immune response, forming the structure of the cell, acting as pumps for the flow of chemicals across cellular membranes, and working as molecular motors (Stryer, 1995). Given their importance, the biotechnology and pharmaceutical sector's focus on proteins is no surprise. The aim of these sectors from an applied standpoint is to develop drugs, vaccines, and diagnostics that target proteins and to develop enzymes for specific industrial processes. Detailed three dimensional structural knowledge is key in the rational design of these products.

The demand for structural information has escalated in the past five years as a result of structural genomics initiatives (Burley *et al.*, 1999). To give an indicator of the amount of structural information that is anticipated, the human genome project, completed in 2003, sequenced an estimated 25 000 genes (Collins *et al.*, 2003). This sequence data all has the potential to be complemented with its three dimensional protein structure. There are three main methods for structure determination including nuclear magnetic resonance (NMR), cryo-electron microscopy (EM), and x-ray crystallography. Of the three, x-ray crystallography is unequivocally the most popular as 86% of the structures deposited in the PDB to date (Feb 2005) were solved using x-ray crystallography (PDB, Berman *et al.*, 2000). It is the most popular method because it has the capabilities of high resolution structural solution, structural solution of large sized proteins, and high-throughput. This last factor is important as to keep apace with the demand for structures, high-throughput structure determination, named structural genomics or structural proteomics, is required. Currently the crystallography community is focusing a great deal of attention on the development and improvement of high-throughput methodologies.

The focus of this thesis is on methodological developments in x-ray crystallography. Chapter two and three of this thesis present methodological improvements that aim to improve the major high throughput crystallography bottlenecks of large scale protein production and crystallization (see section 1.3) (DeLucas *et al.*, 2005). Chapter two presents a new vapor diffusion crystallization protocol, the 'dilution method', which allows for a dramatic reduction in protein consumption as well as increased potential for gaining diffraction quality crystals (Dunlop and Hazes, 2003). Chapter three also presents a new vapor diffusion method, the 'shared reservoir solution' method, that when used in conjunction with a specifically designed crystallization plate may greatly aid in automatic imaging and classification of crystallization trial images (Dunlop and Hazes, 2005). Chapter four investigates the effects of high intensity synchrotron x-ray beams and crystal cryocooling on the final structure (Dunlop and Hazes, 2005).

1.2 Outline of X-Ray Crystallography Process

The x-ray crystallographic process begins with a clone and ends with a final structure. In between, a number of steps are involved including protein expression and purification, protein crystallization, data collection, and structure determination. A generalized outline of the highlights of the x-ray crystallography process is depicted in figure 1.

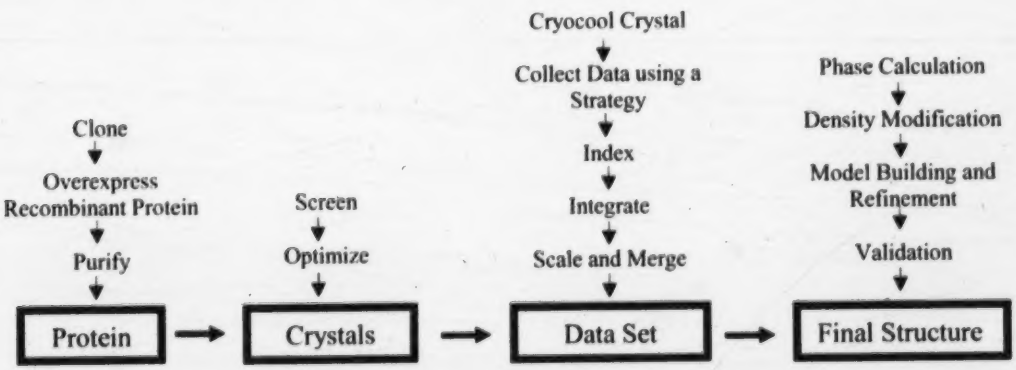

Figure 1.1 Steps in Structure Determination Using X-ray Crystallography.

1.2.1 Protein Production

Obtaining protein for crystallization is the first important stage in x-ray crystallography. A large amount of protein is required, 1-10mg is a good estimate, and it must be very pure. Given the large amount of protein required, protein purification from natural sources is not frequently successful. The relatively new but expensive technique of cell-free protein synthesis shows great promise (Spirin, 2004), however protein production by recombinant methods is the most common (Derewenda, 2004). Protein production using recombinant methods is most popular as it is productive, cost efficient and offers a range of possibilities for modifying the target protein to help in expression, purification, and crystallization (Derewenda, 2004). The first step in obtaining recombinant protein is to clone the target sequence into an expression vector. A frequently used cloning system that allows for the construction of multiple expression vectors in parallel without using restriction endonucleases or ligase is the Gateway® technology (Invitrogen). Following cloning, the recombinant protein is then over-expressed, preferentially in *Escherichia coli,* but also in other bacteria, yeast, or virus-infected mammalian or insect cells. As many constructs do not express well, the ability to construct multiple expression vectors in parallel with the Gateway® technology is a great asset. Following over-expression the protein is purified to homogeneity. Two particularly useful recombinant modifications that assist in purification are the addition of tags such as the hexahistidine tag, or of fusion proteins such as the enzyme glutathione S-transferase (GST). Both tags and fusion proteins can facilitate isolation of the expressed protein by affinity chromatography. The fusion protein also affects expression levels and solubility and needs to be removed prior to crystallization. The general characteristics of a protein preparation that will be successfully crystallized are that it is highly chemically pure, conformationally homogenous, soluble, folded, monodisperse, and stable. Unfortunately, producing sufficient quantities of protein that meet these criteria is a major bottleneck (Rupp, 2003; McPherson, 2004). In chapter two of this thesis, a methodology to reduce protein consumption during the crystallization process, named the 'dilution method' (Dunlop and Hazes, 2003) is presented.

1.2.2 Protein Crystallization

Protein is crystallized so that the diffraction signal of each individual protein molecule is amplified by identical units of protein in the ordered crystal lattice. Crystallization will be described in much greater detail in section 1.3.

1.2.3 Collecting Diffraction Data

Following crystallization, diffraction data are produced by bombarding the crystal with x-rays generated either from a rotating anode or synchrotron source, and then the diffracted rays are collected on a detector. The crystals for this stage may either be used at room temperature, or cryocooled (figure 1.2). Cryocooling crystals is most common, as this has the advantage of protecting them from radiation damage (see section 1.4.2.3). The benefits of room temperature versus cryocooled crystal data collection are discussed in more detail in section 4.3.4.

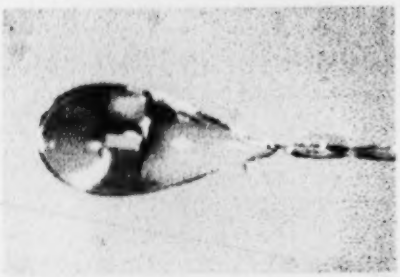

Figure 1. 2 Cryocooled crystal mounted in a loop

Reprinted from *Curr Opin Struct Biol*, 13, Garman. 'Cool' crystals: macromolecular cryocrystallography and radiation damage, 545. Copyright (2003), with permission from Elsevier.

Initially, just a few diffraction images (figure 1.3) are collected so that indexing, the process of determining the symmetry (space group), unit cell parameters, and orientation of the crystal, can be carried out and a data collection strategy determined. This data collection strategy is designed to maximize both the resolution and completeness of the complete native data set that is subsequently collected. Using a synchrotron radiation

source in comparison to a rotating anode x-ray source is advantageous because its x-ray beam is much brighter (see section 1.4.2.1) and the resulting data has improved resolution (see section 1.4.2.2). However, on third generation synchrotrons the beam is so bright that even cryocooled crystals are damaged by radiation. Data collection is therefore, a balance between pushing resolution while minimizing damage (see section 1.4.2.3.3).

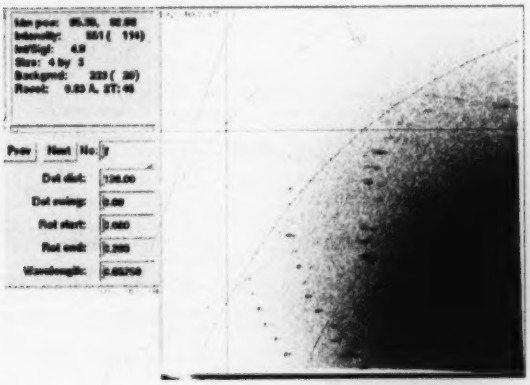

Figure 1. 3 Diffraction image
for the PAK pilin showing diffraction to 0.83 Å. The rings of constant resolution are shown in red.

1.2.4 Structure Solution

All steps from this point onwards use computers and dedicated software. Following data collection the intensity of each diffraction spot is separated from the local background in a process called integration. Because not all data are on the same scale due to changes in x-ray beam intensity, diffracted crystal volume, crystal decay *etcetera*, the difference between symmetry related reflections is minimized in a process named scaling and merging. This is followed by conversion of the intensity information to amplitudes as it is this quantity that is used to calculate electron density maps. However, to do so phase information is also required.

Phase information is lost in the data collection process, and yet phases are necessary for calculating the electron density of the protein. One of a number of strategies is used to solve the phase problem, including Isomorphous Replacement (SIR, MIR, and MIRAS), Anomalous Dispersion (SAD and MAD), and Molecular Replacement (MR). Isomorphous Replacement requires data from a native crystal and one or more crystals with heavy atoms. The intensity differences between pairs of data sets are used to deduce the position of the heavy atoms using Patterson or direct methods, and following on from this the phases are calculated. In comparison to MIR, MAD phasing requires a single heavy atom soaked or seleno-methionine substituted crystal for collection of one or more data sets at specific wavelengths. The small changes in intensities arising from anomalous scattering at different wavelengths are used to calculate the position of the heavy atoms and solve the phases similarly to MIR. The MR technique requires that the structure of a homologous protein is available. Patterson methods are used to obtain the correct orientation and position of the model in the new unit cell and from this phase information is obtained. In all cases, the phase information can be improved using density modification to increase its accuracy. (Taylor, 2003).

Once the phase problem is solved and the amplitudes and phase information is converted into electron density, model building follows (figure 1.4). Model building, which can be automated, involves the interpretation of the electron density map in terms of a set of atomic coordinates. The model is built first by fitting the protein backbone to the electron density, and when the density is sufficiently clear the side chains are fit into the density with the help of sequence information. This preliminary model is then refined against the observed data to improve the fit between model and measurements. As the model improves the phases and corresponding density maps also improve. Model building and refinement are therefore performed in iterative cycles until the crystallographic R-factors, a measure of the agreement between the model and the data, converge.

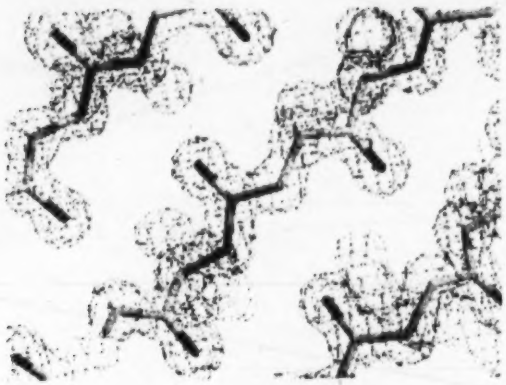

Figure 1. 4 Model building.

Electron density for a β-sheet of the PAK pilin with the main-chain of the protein modeled in.

Validation is the process of assessing the quality of the model. The structure is evaluated on the basis of various criteria. Already mentioned, is the R-factor that measures the global fit of the model to the diffraction data. The R-free factor also does this, but by cross validation. Stereochemical validation investigates whether the torsion angles, bond lengths, bond angles, van der Waals contacts, and residue environment preferences are consistent with what is typical of other proteins. The Ramachandran plot is an important tool here. The programs *PROCHECK*, and *WHAT-IF* provide model based validation analysis. Another important validation tool is the real space correlation coefficient between the calculated model map and the 'experimental' map calculated from observed intensities, areas with low real space correlation often coincide with areas of high B-factors, indicating that model tracing in these areas is likely ambiguous due to lack of electron density. The programs *SFCHECK* (Vagin et al., 1999) and *mapoverlap* from the *CCP4* suite provide real space correlation analysis.

1.3 Crystallization

Protein crystals are of vital importance to x-ray crystallographic structure determination because it is the ordered lattice of proteins in the crystal that acts to amplify the otherwise tiny diffraction signal of single protein molecules. Although there have been amazing increases in crystallization throughput and output as a result of the automation process, crystallization remains one of the biggest bottlenecks in structure determination (figure 1.5) (Bodenstaff *et al.*, 2002; Santarsiero *et al.*, 2002; Chayen, 2003; Rupp and Wang, 2004).

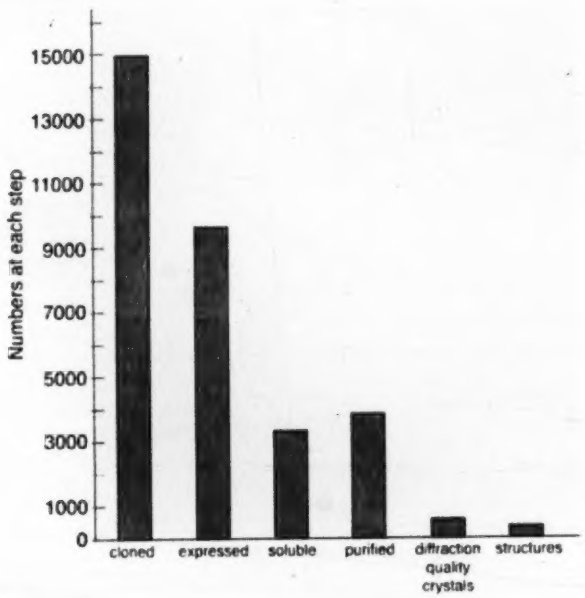

Figure 1. 5 Bottlenecks in structure determination.

Drop off in numbers at each step leading from clone to structure in structural genomics. Chayen, Protein crystallization for genomics: throughput versus output. *J Struct Funct Genomics.* 2003;4(2-3):115-20. With kind permission of Springer Science and Business Media.

Figure 1.5 shows that the step between purified protein and diffracting crystal has the largest percentage-wise drop-off in success. The basic reason for this bottleneck is that the first approach to finding crystallization conditions, screening, has not evolved beyond an empirical trial-and-error approach. The work presented in chapters 2 and 3 of this thesis examines new or little used screening methods. To better understand screening, and crystallization in general, this section will examine the underlying processes of crystallization including nucleation, growth, and cessation of crystal growth. Following on from this, the methods that are currently used to screen and crystallize proteins will be examined, and both current approaches to screening and means by which these approaches may be improved upon to better suit the high throughput environment will be explored.

Discovering conditions in which crystals will grow is usually only the first step in protein crystallization. Optimization of these conditions for optimal crystal size and quality is the very important second step that decides the ultimate success of structure determination. This is because the quality of the final structure is directly determined by crystal size and order (McPherson, 2004b). Despite the importance of this second step the methods used to optimize crystal size and quality are generally quite limited. However, there are a number of successful optimization techniques that are currently used and others that are about to be implemented. The "dilution method", which is presented in chapter 2, is one example of a new method that can be used for optimization (Dunlop and Hazes, 2004). Following on from an introduction to screening, optimization strategies will be examined later in this chapter.

Finally, automation will be explored. Automation of the crystallization process is the means by which throughput has been greatly increased. Despite this success, there are still some stages of automation of the crystallization process, such as automatic imaging and automatic crystal detection that have not realized their full potential. These challenges to automation will be examined.

1.3.1 Supersaturation, Nucleation, Growth, and Growth Cessation

As shown in Figure 1.5, crystallization is a real bottleneck in structure determination. Recognizing this, it is important to increase our understanding of the crystallization process. There are three distinct stages that underlie every successful protein crystallization: nucleation, growth, and growth cessation. Before nucleation can happen the protein solution must become supersaturated (McPherson, 2004b; Chernov, 2003; Asherie, 2004).

1.3.1.1 Supersaturation

Directly prior to setting up a crystallization experiment the protein solution is undersaturated. To obtain crystals, supersaturation of the protein solution must first be induced. Supersaturation is a non-equilibrium state in which the protein concentration is in excess of the solubility limit, but remains nonetheless present in solution (see figure 1.6). Both precipitate and crystals may form from a supersaturated solution because of the thermodynamic driving force in the direction of the lowest free energy. The challenge for the crystallographer is to induce supersaturation of the protein in conditions that will allow crystals to form instead of precipitate.

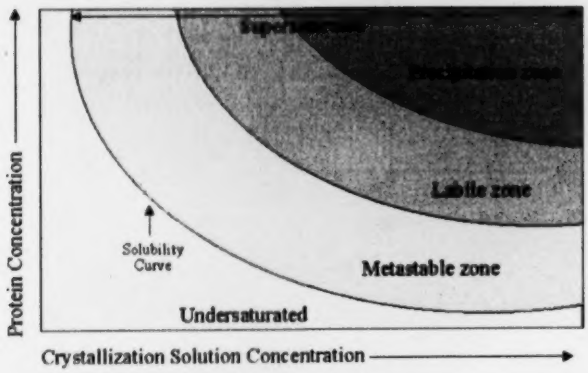

Figure 1. 6 Phase diagram for protein solubility in solution as a function of precipitant and protein concentration.

The phase diagram shows two main regions, the undersaturated and the supersaturated region. These regions are separated by a solubility curve. The solubility curve represents the solubility limit of the protein/precipitant solution, where the crystal/precipitate and solution phases are in equilibrium with one another (saturation). Within the supersaturated region are three zones. Crystals nucleate most favorably in the labile zone and grow most favorably in the metastable zone.

1.3.1.2 Nuclei Formation

Following formation of a supersaturated protein solution the next process in protein crystallization is nuclei formation. There is an activation energy barrier to nuclei formation. This barrier represents the free energy requirement to pass from an initial disordered state of protein in solution to an ordered one, the critical nucleus (figure 1.7). The critical nuclei, as already discussed, are small well-ordered clusters of proteins that form the basis for crystal growth.

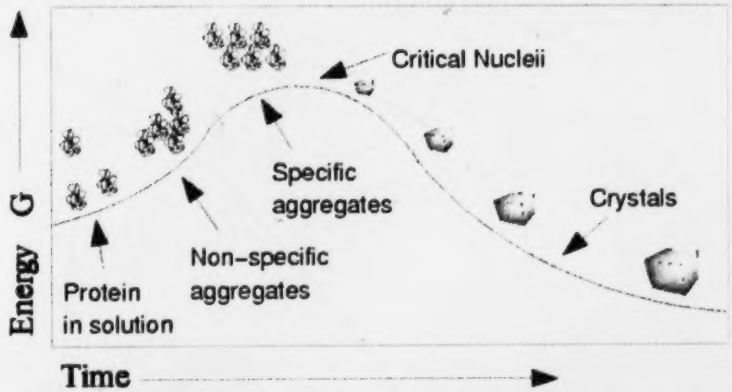

Figure 1. 7 In order for a protein to crystallize it must overcome an energy barrier.

The time axis represents the increasing concentration of the protein solution during the vapour diffusion crystallization experiment before the equilibration point is reached. Figure copyright A.J. McCoy http://www-structmed.cimr.cam.ac.uk.

For a critical nucleus to form it must be thermodynamically driven such that addition of the proteins to the nucleus exceeds their loss. The greater the degree of supersaturation, the greater the thermodynamic driving force to form critical nuclei, and it has been suggested that the supersaturation of the protein must exceed the solubility by a factor of three for successful nucleation (Asherie, 2004). However, if the supersaturation of the protein is too high, disordered aggregates will form faster than nuclei or crystals. The critical nuclei can be theoretically distinguished from specific aggregates and crystals because these clusters of ordered protein molecules have the same probability of growth as of dissolution (Garcia-Ruz, 2003).

The thermodynamic driving force results from specific chemical attraction, in the form of ionic interactions, hydrogen bonds, and van der Waals interactions between protein molecules. In the nucleus, as in the crystal itself, each protein molecule forms at least several intermolecular contacts at each crystal interface with its symmetry related neighbours (Kierzek and Zielenkiewicz, 2001; Chernov, 2003) (figure 1.8). The total surface area from all crystal interface areas on a protein varies widely from protein to protein, and from crystal form to crystal form. For example contact areas may cover as little as 7% to as much as 91% of the whole molecule surface (Chernov, 2003). The more intermolecular contacts, the stronger these contacts, and the greater the total surface area

of these interactions, the greater the stability of the nucleus.

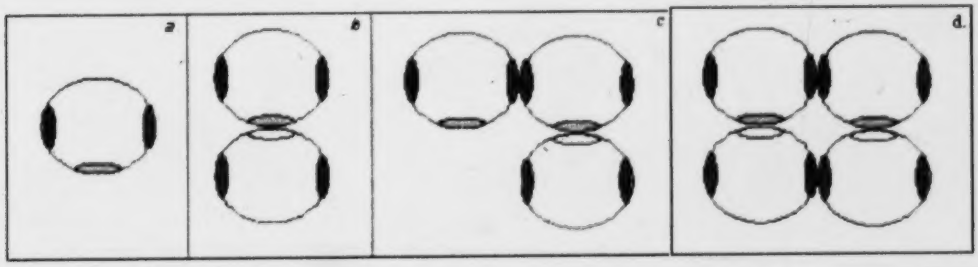

Figure 1. 8 Intermolecular contact areas and nucleus stability.

(a) Single protein molecule. (b) Two protein molecules bonded at single crystal interface. (c) Three protein molecules, each of which is bonded to the neighboring protein(s) at either one or two crystal interfaces. (d) Four protein molecules each of which is bonded to the neighboring proteins at two crystal interfaces. Doubling of the number of protein-protein interfaces helps to stabilize the complex.

1.3.1.3 Crystal Growth

A nucleus forms the seed for subsequent crystal growth. Crystal growth is important because the degree of order of this process has a big influence on the diffraction resolution, mosaic spread, and mechanical stability of the crystals. Crystal growth involves two basic processes, mass transport of protein to the crystal surface, and incorporation of the protein onto the surface (Nanev, 2004).

Transport of protein to the crystal surface happens both by diffusion and convection. Diffusion is the passive process whereby molecules move from areas of high to low concentration. Convective currents are gravity and density-driven and arise from differential density between the high concentration of protein in bulk solution and the low concentration at the crystal-solution interface where protein is depleted by incorporation into the crystal. The effect of convection on crystal growth is to maintain a high level of supersaturation at the crystal interface. As a result, growth never approaches the metastable zone where controlled well ordered growth would occur, but rather, crystals grow very rapidly with a high degree of disorder and defect incorporation. Currently,

convection is under investigation to determine the optimal convective environment for minimal defect incorporation and maximum order of the crystals. A number of techniques, including crystallization at microgravity, and crystallization in gels have been used to stop buoyancy-driven convection. Using these techniques the protein concentration directly around the crystal reaches the metastable zone during crystal growth often resulting in higher diffraction limits, reduced packing defects, larger size, reduced mosaicity, and reduced defect incorporation (McPherson, 1999; Vergara *et al.*, 2003; Judge, 2005). On the other hand, enhanced convection has been shown to reduce the number of nucleations and increase crystal size (Nanev *et al.*, 2004).

Incorporation of protein onto the crystal surface may happen by either of two basic crystal growth processes, crystal face, and tangential layer growth. Crystal face growth refers to the initiation of new layers on the crystal surface (figure 1.9). Recently, atomic force microscopy has been used to visualize crystal face growth mechanisms (figure 1.10). The new layers formed by crystal face growth exhibit step edges to which protein molecules are added by tangential layer growth which results in the extension of new layers (figure 1.9) (Chernov et al., 2003; McPherson *et al.*, 2003).

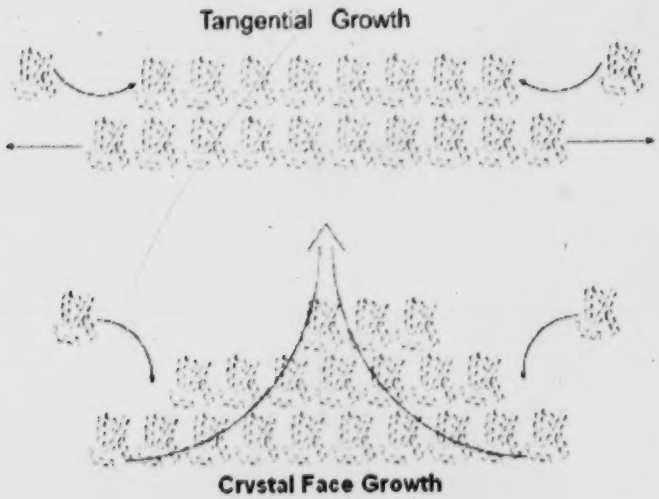

Figure 1. 9 Tangential and crystal face growth

Reprinted from *J Struct Biol*, 142, McPherson, Kuznetsov, Malkin, Plomp. Macromolecular crystal growth as revealed by atomic force microscopy, 32. Copyright 2003, with permission from Elsevier.

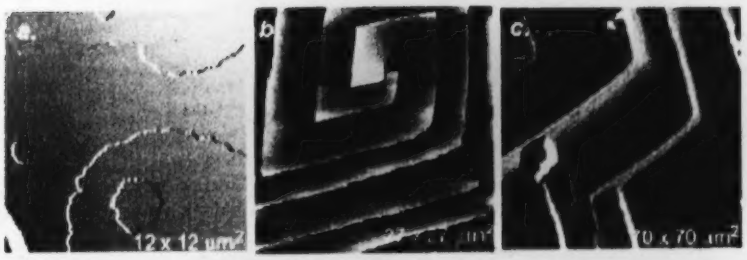

Figure 1. 10 Atomic force microscopy examples of crystal face growth mechanisms.

(a) Growth from 2-D nuclei or islands begins with a two- dimensional nucleation event and once a critical size is reached growth proceeds tangentially from the island. (b) Growth from a screw dislocation happens when steps are continuously propagated in a spiral about a location where either a contaminant has been incorporated or a few molecules misincorporated. (c) Growth from 3-D nuclei involves the formation of multilayer stacks that can be up to a hundred or more layers in height. Reprinted from *J Struct Biol*, 142, McPherson, Kuznetsov, Malkin, Plomp. Macromolecular crystal growth as revealed by atomic force microscopy, 32. Copyright 2003, with permission from Elsevier.

1.3.1.4 Growth Cessation

Crystal growth can end either because the protein solution concentration has reached the solubility limit, which often happens when too many nuclei form, or because the incorporation of impurities has poisoned the crystal surface such that no more protein molecules may be incorporated (McPherson, 1999; McPherson *et al.*, 2003). To be useful for x-ray diffraction, in addition to being well ordered crystals must also generally have dimensions of no less that 10 x 10 x 10 μm. Unfortunately, it is not unusual to obtain large numbers of much smaller crystals or crystalline precipitate. However, all is not lost if the crystal is too small. Crystallographers use seeding to increase crystal size. This involves washing the crystals to remove impurities, then transferring the "clean" crystal to a fresh crystallization drop in which the relative concentration of impurities is low (Bergfors, 2003).

1.3.2 Crystallization Methods

There are a number of crystallization techniques that may be used to produce protein crystals. The goal of all of them is to bring about supersaturation of the solution containing the protein of interest so that well-ordered crystals will form. These methods differ in the parameters manipulated to bring about supersaturation (Durbin and Feher, 1996). The most common methods in use today include the vapor diffusion and batch crystallization methods, which will be discussed in more detail. In addition, free interface diffusion on a microfluidic chip is discussed as it is increasing in popularity.

1.3.2.1 Vapor Diffusion Technique

Vapor diffusion is currently the most popular crystallization method. This method involves equilibration of a crystallization drop against a reservoir solution in a closed system (figure 1.11). The reservoir contains a buffer and precipitating agent, and the crystallization drop solution is typically made by mixing equal volumes of protein and reservoir solution.

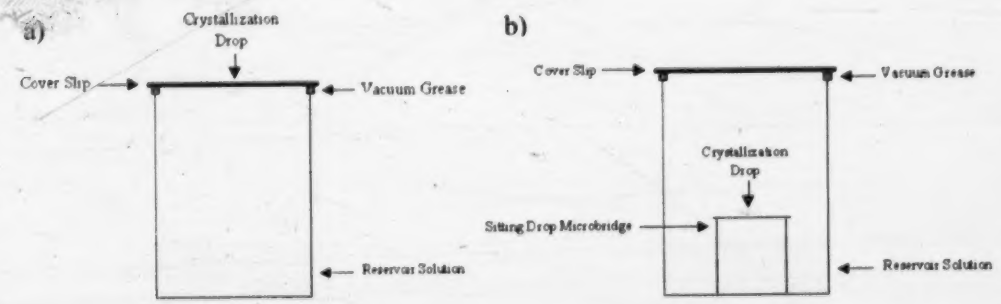

Figure 1.11 A vapor diffusion crystallization experiment.
a) Hanging Drop, b) Sitting Drop.

As a result of the concentration differential between the crystallization drop and reservoir solution, water diffuses through the vapor phase until equilibrium is reached. In an ideal experiment, during the equilibration process the crystallization drop becomes supersaturated and crystals form. This process is illustrated in a phase diagram in figure 1.12.

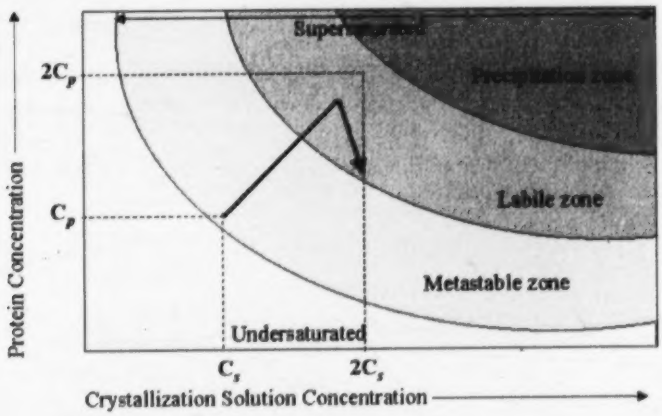

Figure 1. 12 Phase diagram for crystal growth in a vapor diffusion experiment. C_p is the protein concentration, and C_s is the crystallization solution concentration.

Chapters two and three of this thesis discuss advantageous modifications to the traditional vapor diffusion method.

1.3.2.2 Microbatch Crystallization Method

The microbatch crystallization method is used in some of the pioneer high throughput crystallization facilities. In this method, the protein drop and crystallization solution are pipetted under a layer of oil to form the crystallization drop (figure 1.13).

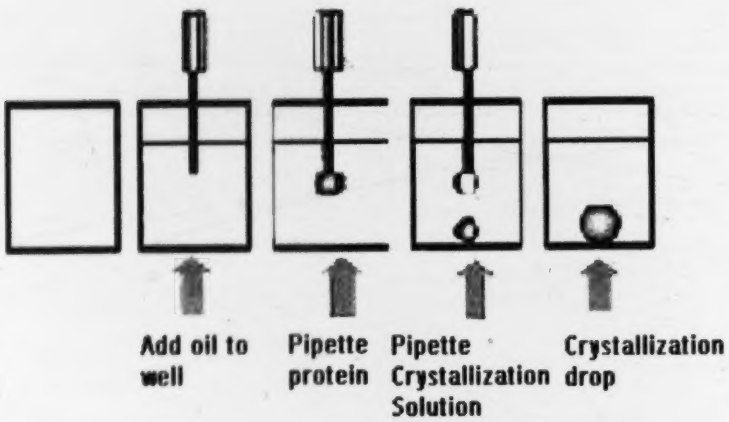

Figure 1. 13 Microbatch method pipetting sequence

Reprinted from *Methods*, 34, D'Arcy, MacSweeney, Haber. Practical aspects of using the microbatch method in screening conditions for protein crystallization, 323. Copyright 2004, with permission from Elsevier.

The major advantages of this method are that there is no problem with evaporation of small size drops during experimental setup, plate sealing is not required, reservoirs are no longer required thus reducing the number of manipulations and reducing the cost, and the oil may protect the protein from oxidation and airborne contaminants (D'Arcy *et al.*, 2004). However, the major disadvantage is that for successful nucleation and crystallization the initial conditions must fall within the labile zone (figure 1.14). This is because the non-water permeable oil used allows no concentration of the drop (Chayen et al., 1992). In comparison with vapor-diffusion, less crystallization space is traversed so the probability of crystallization in any given condition is lower, and this is one of the reasons that microbatch is less popular than vapor diffusion.

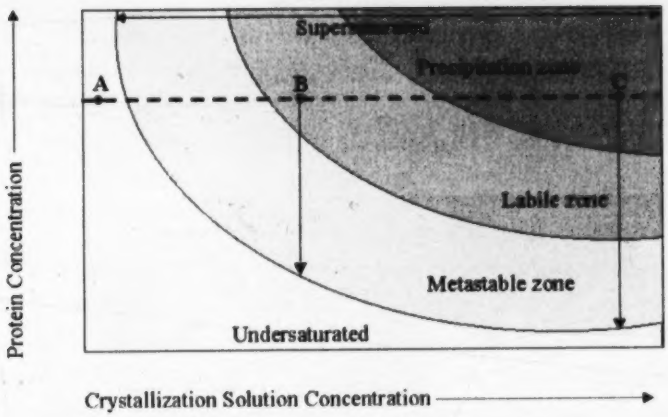

Figure 1. 14 Phase diagram for microbatch crystallization experiments.

(A) Initial conditions are in the undersaturated zone and the drop will stay clear, (B) initial conditions are in the labile zone and in this instance result in nucleation and crystal growth, and (C) initial conditions are in the precipitation zone so the protein precipitates.

Today it is agreed that this method is best suited for optimization of established conditions (D'Arcy, *et al.*, 2004). For crystallization screening, the "modified microbatch" or "microbatch diffusion" type experiment is recommended. In this modification, a water permeable oil such as Al's Oil (Hampton Research) is used enabling the drop to concentrate over time, and thereby traverse more of crystallization space. The disadvantages to using this modified method is that there is no end point to the final concentration of the crystallization drop so that drops must be watched carefully over time. Then, when the crystals have stopped growing the Al's oil must be replaced with paraffin oil to stabilize the crystals. In addition, it can be difficult to harvest the crystals through oil, and determination of cryosolution concentration is difficult, as it requires accurate knowledge of the crystallization solution concentration at the stabilization point (D'Arcy, *et al.*, 2004).

1.3.2.3 Free Interface Diffusion and Microfluidics

In the Free Interface Diffusion method (FID), protein and reagents combine gradually by diffusion under the influence of a concentration gradient (figure 1.15).

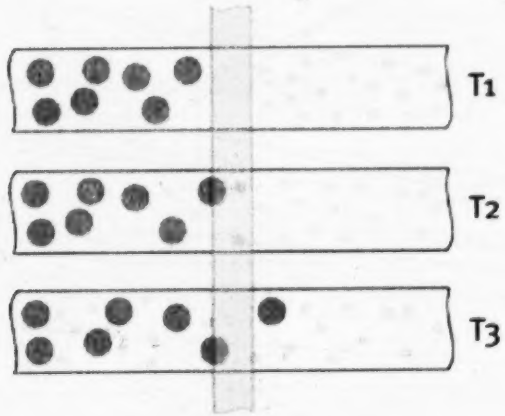

Figure 1. 15 FID crystallization.

The effect of diffusion on mixing protein (large spheres) and reagent (small spheres) at three successive time points (T1, T2, T3). Copyright Fluidigm http://www.fluidigm.com/products.htm.

Along the concentration gradient precipitate, crystals of different sizes, and clear solution are simultaneously present. The advantage of FID is that a larger region of crystallization space is traversed than when using any other crystallization method (figure 1.16), and therefore the likelihood of crystallization success is also higher (Chayen, 2004).

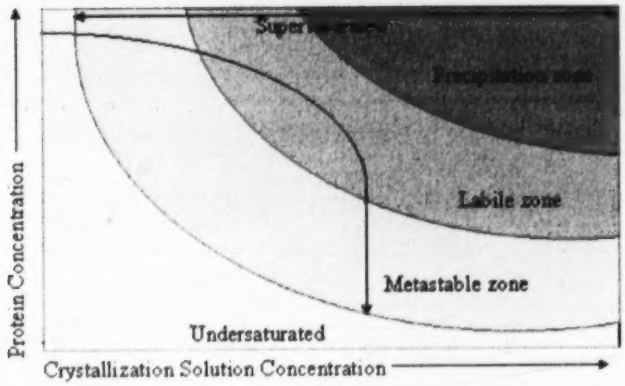

Figure 1. 16 Phase diagram for free interface diffusion.

The phase diagram is for one point in the chamber that initially only contains the protein solution.

FID has recently undergone a resurge in popularity as a result of its application with microfluidic chips. Microfluidic FID uses high density screening chips (figure 1.17) to create a nanoscale (reagent volume 2.55nL, protein volume 0.75nL) high throughput form of FID. Its other advantages includes that its volumes are more precise, and its downstream image analysis is more accurate than when using any other method (Hansen and Quake, 2003). Its disadvantages are that the chip price is too expensive to fall within the grasp of many crystallographers, there is no hope of harvesting the crystals, and reproducing the results is somewhat questionable.

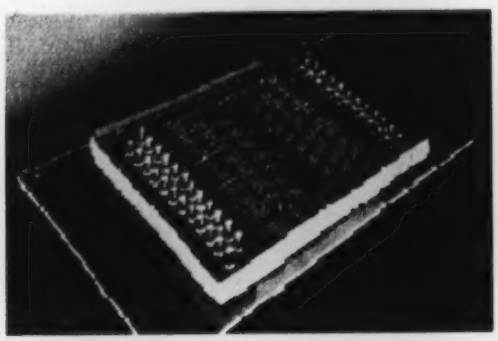

Figure 1. 17 A microfluidic chip

Reprinted from *Curr Opin Struct Biol*, 13. Hansen and Quake. Microfluidics in structural biology: smaller, faster em leader better, 538. Copyright (2003), with permission from Elsevier.

1.3.3 Screening

Thus far, a method for the *de novo* prediction of crystallization conditions for a previously uncrystallized protein has not been developed and is unlikely any time soon (Luft, *et al.*, 2003). The essence of the current approach to crystallization is basically trial and error, and has remained unchanged for the past 100 years. This approach involves two steps, screening followed by optimization. Screening tests whether the protein of interest will crystallize in any one of a gamut of crystallization solutions usually purchased in screening kits. The individual members of the crystallization screen all vary from one another according to their particular combination of precipitant, additives, buffer and pH. The idea behind screening is to traverse as much of 'crystallization space' as practical in order to maximize the chance of finding successful crystallization conditions given the amount of protein, time, and resources available. Crystallization space is multi-dimensional and includes all parameters, physical, chemical, and biochemical that effect crystal formation. However, it is generally only the chemical parameters plus sometimes temperature that are varied during screening. A full list of crystallization parameters is given (table 1.1). All of these parameters differ in their relative effects on crystallization success (Durbin and Feher, 1996). Section 1.3.3 will explore the different types of crystallization screen, and the number of conditions that can be tested.

Table 1.1 Parameters that effect crystallization.

Reproduced with modifications from *Methods*, 34, McPherson. Introduction to protein crystallization, 254. Copyright 2004, with permission from Elsevier.

Physical	Chemical	Biochemical
Temperature	PH	Macromolecule purity
Surfaces	Precipitant type	Macromolecule size
Methodology	Precipitant concentration	Aggregation state
Gravity	Ionic strength	Post-translational modifications
Pressure	Specific ions	Source of macromolecule
Time	Degree of supersaturation	Proteolysis/hydrolysis
Vibrations	Redox environment	Chemical modifications
Electrostaic/magnetic field	Macromolecule conc.	Genetic modifications
Viscosity of the medium	Metal ions	Macromolecule symmetry
Rate of equilibrium	Croslinkers/polyions	Macromolecule stability
Nucleants	Detergents/amphophiles	Isoelectric point
	Impurities	Sample history

Screening efficiency has been the focus of considerable research. This is because the amount of protein available is usually a limiting factor and because setting up and monitoring crystallization experiments is very labor intensive. Both the content of and logic behind the screen types and the number of conditions tried per protein are subjects of interest.

The choice of screening conditions must be comprehensive so that crystallization space is widely sampled, and yet efficient so that crystallization success is achieved in the least number of trials possible. A full factorial screen (figure 1.18 A) would be all encompassing and do an exhaustive fine screen of all known parameters. However, as the number of conditions required would explode, this type of screen is not practical (Luft *et al.*, 2003). However, there are several types of screening methods that hope to achieve both comprehensive and efficient sampling, including the incomplete factorial, sparse matrix, and random screens. Brief descriptions of each follow.

Incomplete factorial screening (figure 1.18 B) samples a limited subset of possible combinations of parameters in a systematic balanced fashion using no prior knowledge of

possible parameter success. Sampling of parameter space gives a balanced representation of all parameters and is efficient (Carter and Carter, 1979; Segelke, 2001; Rupp and Wang, 2004).

The random screen selects parameters to be tested in a purely random manner (figure 1.18 C). This is considered by Segelke to be the most efficient screen type especially when success rates are low or clustered (Segelke, 2001). Random screens can be computer generated. For example, CRYSTOOL is a computer program that uses random number generation to choose each component in each experiment of a screen. Each component is assigned a particular frequency, and the probability of choosing a particular component is determined by this frequency (Segelke, 2001).

Sparse matrix screening (figure 1.18 D) selects reagents based on prior successes with other proteins. These reagents are grouped into classes such as precipitant, buffer, and additive and a limited number of combinations of reagents are selected (Jancarik and Kim, 1991). Many of the commercially available screening kits, such as Crystal Screen (Hampton Research), and Wizard Screens I and II (Emerald BioStructures), are sparse matrix screens. The advantage of this method is that only crystallization conditions that are tried and true are being tested. The drawback is that repeated use of these commercial screens results in an over sampling of these areas of crystallization space when there may be other potential hotspots that have never been tried (Page and Stevens, 2004; Rupp and Wang, 2004). In addition, many of the conditions used in any one screen may be redundant (Gao *et al.*, 2005).

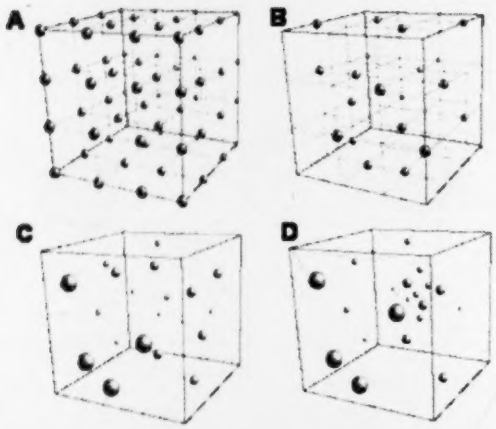

Figure 1. 18 Screening Methods.

(A) Full factorial, (B) Incomplete factorial, (C) Random, (D) Sparse Matrix. Figures copyright A.J. McCoy http://www-structmed.cimr.cam.ac.uk.

Irrespective of the type of screen used, the number of screening experiments to conduct on a protein of interest is an important consideration given the limited amount of protein available, the time and expense it takes to set up and monitor the experiments, and the limited likelihood of crystallization success for any given protein. Investigations have revealed that the majority of soluble proteins ~65% crystallize rarely or not at all, and the likelihood that a soluble protein has the potential to form crystals given current crystallization techniques is around 35%. For the 35% of proteins that will crystallize, the number of screening conditions that yield crystals ranges anywhere from 0.2 % to 15% (figure 1.19) (Segelke, 2000).

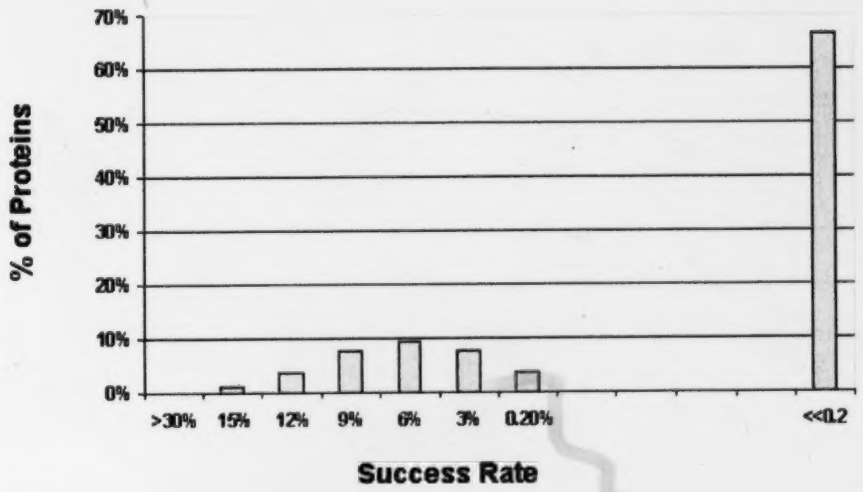

Figure 1. 19 The success rate of crystallization

This graph shows the % of proteins (y axis) that will crystallize in <<0.2 - >30 % of all crystallization conditions tried (x axis). Segelke, B.W. (2000). *ACA Meeting Series.* 27, 43. Reproduced with kind permission from the International Union of Crystallography.

Given the odds, in many respects the most strategic method of crystallization is to carry out a few hundred randomly generated or sparse matrix crystallization trials, and then if there were no successes, to work on either increasing the purity of the protein sample or on protein engineering to create a more readily crystallizable protein (see later this section) (Segelke et al., 2004). However, the minimum number of trials is not always the most effective. Sampling more conditions is likely to produce more lead conditions. As crystal grown in different conditions may have different space groups some of which are easier to optimize, and some conditions may yield crystals that are better ordered and diffract more strongly than others, there are advantages to trying more than the minimum number of conditions (Luft *et al.*, 2003). Some crystallographers even feel that given the high throughput capabilities of crystallization robots and imagers, and the relatively higher cost of producing protein than of carrying through with crystallization trials, that the cost of missing a rare but viable crystallization condition outweighs the inefficiencies of broader screening (Bard, *et al.*, 2004).

The greatest reduction in the number of crystallization conditions screened would be achieved if crystallization were able to move from a trial-and-error to a prediction based approach. It is hoped that the current high-throughput initiatives will generate and preserve the crystallization information that is required as a database that may then be mined for predictive information. Assembling crystallization databases requires the recording of crystallization outcomes both positive and negative, and including information on reagents, physical factors, methodological approaches used, and the biochemical and biophysical history of the protein itself. This is currently underway in the NIH funded Structural Genomic Centers. Also in current development are means to analyze these crystallization databases. This analysis process named data mining, as yet in its infancy, uses both statistical and machine learning approaches based on neural networks (McPherson, 2004a; Rupp and Wang, 2004) or decision trees (Berlone, *et al.*, 2001).

There are three general levels of data base analysis. The most basic is a global success analysis, describing which conditions work best for all proteins (Rupp and Wang, 2004). This has in fact been undertaken by a number of structural genomics consortia who have come up with minimal sparse matrix screens containing between 6 and 30 conditions that yield crystallization success with the majority of their crystallizable proteins (Page and Stevens, 2004; Gao et al., 2005). At the next level special screening kits may be derived for specific classes of protein. These screens have already been developed for membrane proteins (Sigma-Aldrich, Hampton Research, Jena Biosciences, Nextal, UCLA Crystallization Workshop, June 21, 1993), and DNA binding proteins (Hampton Research, Scott *et al.*, 1995). Ultimately, the conditions for crystallization of specific proteins might be predicted based on prior knowledge about the sample derived from the sequence such as monomer weight, pI, affinity tag, source, cellular location, and protein interactions (Rupp and Wang, 2004).

Although it is true that for a crystal screen to be most efficient is must use the most efficient sampling protocol and/or sample from a region of crystallization space that has

the greatest likelihood of success (Segelke, 2001), given the low probability of crystallization success for many proteins it may be that the most important parameter is the protein (Segelke, 2001; Dale, 2003). By altering the intrinsic properties of a protein that does not readily crystallize it has been possible to increase its probability of crystallization success. One option is to consider homologous proteins from another species. Often the homologues will differ by several residues and have greatly different crystallization potential. But in addition to using homologues, there are at least six types of engineered modifications that may be used to increase the likelihood of crystallization success. (1) The most common modification is truncation of highly flexible or unfolded segments, particularly at the N and C termini, to decrease the entropy penalty of crystal formation. Entire domains may even be deleted. (2) Selected hydrophobic surface residues may be mutated to enhance protein solubility for those proteins that precipitate at the concentrations that are required for crystallization. (3) For proteins expressed in eukaryote hosts glycosylation sites may be removed to increase homogeneity and reduce flexibility. (4) Exposed sulfhydryl groups may be mutated to serines if they form non-native disulphide bonds and lead to sample hererogeneity or aggregation. (5) Easily crystallizable fusion proteins, such as maltose binding protein, may be added to act as a scaffold to increase crystallizability of the protein of interest. (6) Point mutations of surface residues with high conformational entropy, specifically lysines and glutamates, may increase the surface area of the protein that is able to make crystal contacts and therefore increase the likelihood of crystallization success (Dale *et al.*, 2003; Derewenda, 2004). (7) Reductive methylation of free amino groups including lysine residues and the N-terminus of the protein may greatly improve crystal quality (Rayment, 1997).

1.3.4 Optimization

Crystals must meet certain criteria to be useful. Typically they must diffract to better than 3 Å resolution which means that they must be well ordered and they must be of sufficiently large size, generally at least about ~10 μm in each dimension. However, as the crystals obtained during initial screening frequently do not meet these criteria, the crystallization conditions must be optimized to obtain bigger and better ordered crystals.

Optimization is a multiple step procedure with every step depending on the outcome of the previous one. Typically, the first stage of optimization is to fine-tune the parameters such as pH, and precipitant concentration around the initial hit condition using a grid screen format. This requires the production of tailor made solutions. If this fails, some of the original ingredients may be replaced, arbitrary reagents may be added, and experimental set up factors such as drop size and protein to crystallization solution ratio may be modified. If still further optimization is required, which it frequently is, there are many other techniques that may be used, however, aside from seeding very few of these techniques are ever employed. These techniques fall into two categories, optimization by active control of the crystallization process, and optimization by decoupling nucleation and growth. Examples of optimization by active control of the crystallization process include, crystallization in gels to reduce convection (see section 1.3.4) and slow down the equilibration kinetics to favor the formation of fewer better ordered crystals (Chayen, 2003). Crystallizing between two layers of oil to reduce excess nucleations on the crystallization drop well walls resulting in larger crystals (Chayen, 2003). Crystallization at high pressure (Lin *et al.*, 2005; Kadri, *et al.*, 2005) or at microgravity (Judge et al., 2005) to increase crystal order. Optimization by decoupling nucleation and growth is possible because these two processes optimally require different conditions. The most common method to decouple nucleation and growth is seeding. Seeding uses previously nucleated crystals either whole or crushed, and introduces them into new drops equilibrated at lower levels of saturation (Bergfors, 2003). Another way to decouple nucleation and growth is to quench nucleation using dilution. This is achieved by incubating until a minimal number of nucleations have occurred, then after a specific time lag, diluting the reservoir solution to produce conditions more suitable for crystal growth (Chayen and Saridakis, 2002).

1.4 Methods Improvements in Crystallography

Since all of the work described within this thesis has a strong connection with methods improvements in crystallography, be it the development of a new method or research on the effects of particular methods on the final structure, a brief description of some of the

most powerful methods improvements that relate to the work in this thesis are given in the following sections. These relevant methods improvements include: automation of the crystallization process for high-throughput, crystal cryocooling to prevent radiation damage, and the increase in brilliance of synchrotron radiation sources.

1.4.1 Methods Improvements in Crystallization

1.4.1.1 Automation of the Crystallization Process

Automation of the crystallization process was begun in the late 80's and early 90's. The progress of automation has been rapid, and is now commonplace in commercial labs and, dependent on cost, is being used in academic labs also. It has the capability to greatly increase throughput while decreasing the manual labor requirement. Automation is possible at all stages of the crystallization process including crystallization solution preparation, crystal experimental plate setup, plate sealing, delivery of plates to storage, delivery of plates to imaging, imaging, automatic classification, and record keeping. In some high throughput commercial labs automation is used at all stages, however, whereas some stages of the automation process are well developed and run smoothly, other stages are still being developed. In addition, given the cost of automation processes, some labs choose to automate only a few crucial stages such as crystallization set-up and automatic imaging, resulting in much higher throughput and large time savings. Whereas automation robots were initially developed with the idea of automating classical crystallization methods, it is likely that it is more strategic to adapt the crystallization method with a robot in mind. The two methods development papers included as chapters two and three of this thesis introduce modifications of traditional methods that do just this. These new innovations are strongly linked to the automation process, coming at it from the perspective of increasing crystallization success by improving crystallization set up, automated optimization, automatic imaging, and classification. These stages of the automation process will be examined in more detail.

1.4.1.2 Crystallization Trays

Vapor diffusion sitting drop crystallization plates for automated set up will be discussed here because chapter three presents a novel high density drop plate designed for this purpose (Dunlop and Hazes, 2005). Plates used for automated crystallization experiments generally conform to dimensions set out by the Society of Biomolecular Screening (SBS). These plates usually have 96 wells, and have been designed for high-throughput drop set up. Examples of plate models include the Greiner round and square well CrystalQuick™ plates and the Corning Crystal*EX*™ plates. The differences between the plates involve the shape of the sitting drop shelf design, which may be a flat or curved bottom well, and the material the plate is made from, which may be either polypropylene, polystyrene, or an 'advanced optical polymer'. These plates work well for crystallization set-up, however the downstream image capture step is substandard (Hui and Edwards, 2003). This is a result of two problems with image capture, firstly, of drops adhering to the sides of the wells and, secondly, low optical clarity of the plate. A novel solution to all these problems is presented in chapter three (Dunlop and Hazes, 2005).

1.4.1.3 Crystallization Robots

Automation improves the efficiency of crystallization setup by both reducing the amount of protein required for each screen and by decreasing manual labor (Santarsiero, *et al.*, 2002; Bard *et al.*, 2004). There are many crystallization robots on the market. They vary significantly in their: number of heads, minimum and maximum size of drop, flexibility, time savings, accuracy, and also in the stages of crystallization set up in which they are able to participate. The most flexible of these are the SolutionMaker from DeCode Genetics (Bard *et al.*, 2004), the Screenmaker 96+8™ Xtal from Innovadyne (www.innovadyne.com), and the Honeybee from Genomic Solutions. The latter has been shown to be able to generate custom screens, dispense premade screens, and set up the nanovolume crystallization experiment (Hazes and Price, 2005).

The advantages of crystallization robots are huge. They include: (1) the amount of protein

required per experiment is reduced because automated dispensers can accurately dispense smaller volumes than can be dispensed by hand. For instance 0.5-1µl is the limit of accurate manual dispensing, whereas some robots can dispense volumes as small as 20 nl (Santarsiero *et al.*, 2002). This leads to large savings in protein consumption, an enormous advantage, because producing adequate amounts of protein for crystallization is one of the bottlenecks in crystallography (DeLucas *et al.*, 2005). (2) Increased throughput per unit of time and decreased manual labor requirements. Many crystallization robots are able to set up in the order of 6000 – 100 000 crystallization trials per day (Bern *et al.*, 2004). It would take months to years to set up the same number of experiments by hand. (3) Savings in the volume of reagent used per experiment. This is particularly important when commercially purchased crystallization kit solutions are being used. The paper presented in chapter three of this thesis (Dunlop and Hazes, 2005) enables even greater reagent savings. (4) Space savings. The 96 well plates are both smaller than the manual set up linbro plates, and fit four times as many drops. This space saving is very important because high throughput labs must store a large number of crystallization plates. Additional possibilities for space saving, resulting from the development of high drop density crystallization plates, are addressed in chapter three of this thesis (Dunlop and Hazes, 2004).

There are two challenges in the use of crystallization robots. The first is accurate delivery of nanoliter volumes of solutions with widely varying viscosities (DeLucas *et al.*, 2005). This challenge is addressed in the second chapter of this thesis, titled "When less is more: a more efficient vapor-diffusion protocol" (Dunlop and Hazes, 2003). The second challenge in the use of crystallization robots is in minimizing water loss during experimental setup. This problem can be virtually eliminated if the microbatch under oil crystallization technique is used (DeLucas *et al.*, 2005). However, most crystallographers prefer to use the vapor diffusion method. In this case using a humidity chamber to retard evaporation is a big help. The "dilution method" (Chapter 2, Dunlop and Hazes, 2003) also provides additional benefits in minimizing disruption and uncertainty to the crystallization experiment through water loss.

The process of optimization is not easy to automate since it must be adapted to each case individually. As a result, many high-throughput institutes only use diffraction quality crystals that form during the screening process, leaving numerous protein structures unsolved (Chayen, 2003). This strategy has shifted the crystallization bottleneck from the screening to the optimization stage (Leulliot *et al.*, 2005). However, very recently, a high-throughput facility, the Yeast Structural Genomics Project, has implemented some simple optimization procedures to make a standard one size fits all optimization process. The optimization procedures used include, fine tuning around the initial screen success by changing precipitant concentration and buffer pH, additive screening, adding an oil permeability barrier to slow down equilibration, and crystallization in gels (Leulliot et al., 2005). Another optimization process that can be easily implemented by high throughput operations is presented in chapter 2 of this thesis.

Automated setup allows thousands of experiments to be setup per day. The results of these experiments must then be followed at approximately eight to ten time points in the subsequent months. This creates an overwhelming number of drops to be viewed. Fortunately automated imaging makes the process faster and more comfortable.

1.4.1.4 Automatic Imaging

Automatic imaging robots rapidly record digital images of crystallization experiments, and store them in a database for subsequent analysis. The advantages of this process include: (1) Speed. The crystallographer is able to scan a many images and score them for crystallization success at a much faster pace and more comfortably than when using a manual microscope. (2) There is a permanent record of crystallization results, so any changes over time can be examined. (3) There is the potential for automated scoring (see section 1.4.1.5).

Initially the quality of the images was not as clear as when using a manual microscope, and although recent developments have improved image clarity, problems remain. The causes of the lack of clarity and the improvements, if relevant, include: (1) limited depth

of field. The solution to this has been to take several images at successive focal depths and compose them into a single image (DataCentric Automation). (2) Shadow effects at the edge of drops. As solutions, improved illumination, including use of a LED array that allows spatial variation in light intensity (Bard, *et al.*, 2004) and a computational approach termed "sweet image" (RoboDesign). (3) The optical quality of the plastic used in plate design makes viewing difficult. (4) The crystallization drops frequently, as a result of electrostatics and capillary action, cling to the edges of the well resulting in the obscuration of the drop from the field of view. Chapter three of this thesis addresses problem 3 and 4 above with the development of a novel crystallization plate (Dunlop and Hazes, 2005).

1.4.1.5 Automatic Classification

Even when an automatic imager is used the huge number of crystallization experiments generated by crystallization robots overwhelms the capability for manual observation and scoring of these experiments (Pusey *et al.*, 2005). Technologies for rapid, accurate evaluation of crystallization results have been developed but have not yet matured sufficiently to be in general use (Wilson, 2002). This is largely due to the variability present in images of crystals and precipitates, making automatic classification a very challenging machine learning problem (Bern, *et al.*, 2004). In addition, crystallization plates present challenges for classification as a result of the flow lines that are introduced into the plastic during the manufacturing process, and the electrostatic properties of the plate.

There are a few commercially available classification systems and others under development or used privately, all of which vary in their complexity. For the least complex, the system offered by Formulatrix scores whether the drop is clear or not (www.formulatrix.com). Elimination of clear drops from the manual scoring greatly reduces the human workload, but still leaves many drops for human scoring. More complex systems such as RoboDesign CPXO Automatic Scoring software, and the Wilson (2002) classification system differentiate the content of the drops into a number

of categories. For example the RoboDesign software scores as clear, precipitate, crystal, or other (http://www.robodesign.com/cpxo.shtml), and the Wilson (2002) system scores as crystals, aggregated crystals, interesting objects, skin or junk. In order to make these classifications, the process generally involves first, identification of the crystallization drop, its boundaries and the crystals themselves using edge detection. Edge detection is currently the most common method for crystal identification. It primarily identifies straight lines as this is the key feature that sets crystals apart from other objects. However, the false negative error rate in crystal identification is known to be very high. For the (Wilson, 2002) system the false negative error rate is 13% for crystal identification and 23% for non-crystal objects. This is regrettable, as this type of error rate is too high for automatic classification to replace manual inspection (Walter, *et al.*, 2005). A further disadvantage of the systems is that they cannot evaluate crystalline or quasi-crystalline results, which can be optimized to obtain crystals (Pusey, 2005).

To overcome these difficulties, the property of birefringence and the rotating-polarizer microscope technique would be an attractive addition to the use of edge detection for crystal detection (Echalier *et al.*, 2004). Protein crystals because of their anisotropic nature are birefringent. Basically, when placed between cross polarizers they will transmit different colours depending on the angle of the polarizers. The rotating-polarizer microscope technique uses monochromatic light, a rotating polarizer and a circular analyzer (http://www.metripol.com), and is more sensitive than the traditional birefringence method (Echalier *et al.*, 2004). An advantage of using birefringence as a crystal detection method is that it is simpler than edge detection, with only the changes in transmitted light need be monitored (Bodenstaff *et al.*, 2002). In addition, it is sufficiently sensitive to detect microcrystals in a precipitate that appears amorphous using traditional inspection methods, thus increasing the region in the phase diagram in which crystals can be detected (Echalier *et al.*, 2004). Disadvantages include that it is unable to detect crystals with the cubic space group (Bodenstaff *et al.*, 2002). In addition, foreign birefringent material will affect detection. One issue that needs to be considered for future use of this technique is the requirement for inexpensive commercially available birefringence-free crystallization plates (Echalier *et al.*, 2004). Chapter three of this thesis

introduces a non-polarizing crystallization plate that if used with a glass lid and commercially produced would help make birefringence viable (Dunlop and Hazes, 2005).

1.4.2 Methods Improvements in Data Collection

1.4.2.1 Increase in Brilliance of Synchrotrons

Synchrotron beam brilliance has increased by many orders of magnitude in the past 20 years, resulting in the advantages of reduced data collection times and increased resolution. To document this historical increase in brightness, a comparison of rotating anode x-ray source, first, second, and third generation synchrotron beams shows values of 10^9, 10^{12}, 10^{15} and 10^{18} photons s^{-1} mm mrad $^{-2}$, respectively (Bergmann, personal communication). The brightest sources, 3rd generation synchrotrons, were developed in the 1990's. The major technological difference between these and their 2nd generation counterparts is the addition of undulators to the 3rd generation synchrotron (Helliwell, 1998). An undulator is comprised of an array of approximately 100 magnets with alternating poles and is inserted into the straight section of the synchrotron storage ring. It enhances the output of synchrotron radiation through constructive interference and the resulting beam has high flux, highly parallel X-rays, and is focused into a small cross-sectional area (Hendrickson, 2000). The increased brilliance of X-ray sources has many advantages. As already mentioned, data sets are collected much more rapidly using very short exposure times, and the high resolution weak reflections may be recorded (see chapter 4) (Pernot-Rejmánková, 2002). However, along with these positive factors one drawback of the increased brilliance is radiation damage (see section 1.4.2.3). Chapter 4 of this thesis examines the positive effect of a brilliant synchrotron source, i.e., increased resolution, as well as at the negative effect of radiation damage on the final structure (Dunlop and Hazes, 2004). These effects will be examined in more detail in the following sections.

1.4.2.2 Ultra-high Resolution Structures

The ultimate goal of structural biology is to understand protein structure and function in the greatest detail possible. For a handful of proteins it has been possible to obtain sub-atomic information, including crambin determined to 0.54 Å resolution (Jelsch et al, 2000) and aldose reductase solved to 0.66 Å resolution (figure 1.20) (Lecomte *et al.*, 2004). The sub-atomic information available in these structures includes valence electron density distributions, protonation states, and deviations from standard geometry (Lecomte *et al.*, 2004).

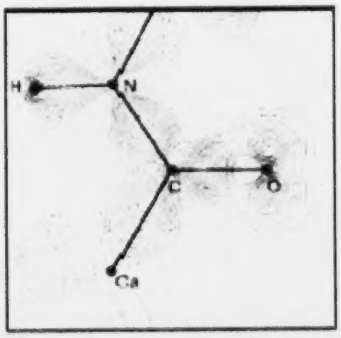

Figure 1. 20 Static deformation density of a peptide plane in crambin

Reprinted from Proc Natl Acad Sci U S A. 2000 Mar 28;97(7):3171-6. Jelsh, Teeter, Lamzin, Pichon-Pesme, Blessing, Lecomte. Accurate protein crystallography at ultra-high resolution: valence electron distribution in crambin. Copyright 2000 National Academy of Sciences, U.S.A.

Following on from the subatomic structures, the next group of structures that reveal protein detail at only a slightly lower level are the ultra-high resolution structures. The PAK pilin structure presented in chapter four falls into this category (Dunlop and Hazes, 2004). Therefore, generalizations of ultra-high resolution structures will be discussed in more detail here. Ultra-high resolution structures are those with a resolution better than 1.0 Å. Structures with a resolution that shows this level of detail are rare. Factors that enable such high resolution include both features of the instrumentation and techniques

used to collect the data, including use of high brilliance synchrotrons, use of very short wavelength x-rays, and crystal cryocooling, and features of the protein and crystal themselves, including small protein size (Howard, *et al.*, 2004), well ordered crystals, and low solvent content (less than 35% v/v) (Lecomte, *et al.*, 2004).

The increased detail that ultra high resolution structures show as compared to lesser resolution structures generally includes several features. (1) Hydrogen atom electron density can be visualized (figure 1.21) (Kuhn *et al.*, 1998; Howard *et al.*, 2004; Dunlop and Hazes, 2005). The presence of electron density for specific hydrogens depends on the temperature factor (B-factor) of the heavy atom to which the hydrogen is bonded (Howard *et al.*, 2004). (2) At ultra-high resolution more alternate conformations are often resolved in electron density maps (Betzel *et al.*, 2001) (figure 1.22), and the occupancy factors corresponding to the relative residence times of each conformation can be assigned (Declercq *et al.*, 1999; MacArthur and Thornton.,1999; Howard *et al.*, 2004). (3) Carbon, nitrogen, and oxygen atoms can be differentiated based on electron density (figure 1.23) (Kuhn *et al.*, 1998; Betzel et al., 2001).

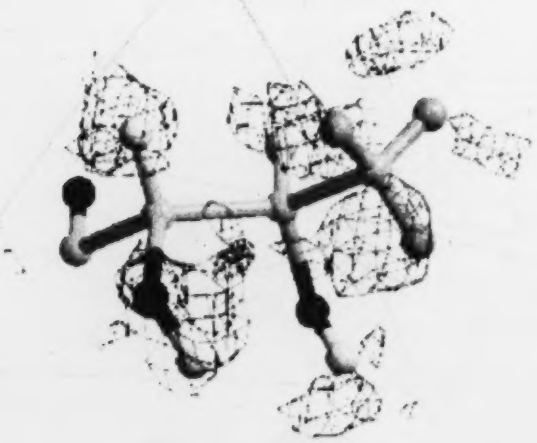

Figure 1. 21 Hydrogen atom density.

H atom density on Thr 95 of a pilin structure determined at 0.78 Å resolution. The $F_o - F_c$ difference map, from the model before H atoms were included, is contoured at 1.5 σ. Published in Acta Crystallogr D Biol Crystallogr. 2005 Jan;61(Pt 1):80-7. Dunlop, K.V., Irvin, R., and Hazes, B. Reproduced with kind permission from the International Union of Crystallography.

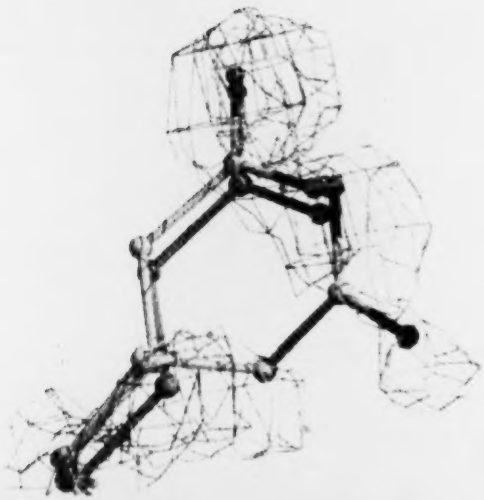

Figure 1. 22 Alternate conformations

$2F_o - F_c$ map for the alternate conformations of the Glu 27 side-chain in the 1.51 Å resolution PAK pilin structure collected from cryocooled crystals (yellow) and for the single conformation collected from room temperature crystals to 1.63 Å resolution (green). Published in Acta Crystallogr D Biol Crystallogr. 2005 Jan;61(Pt 1):80-7. Dunlop, K.V., Irvin, R., and Hazes, B. Reproduced with kind permission from the International Union of Crystallography.

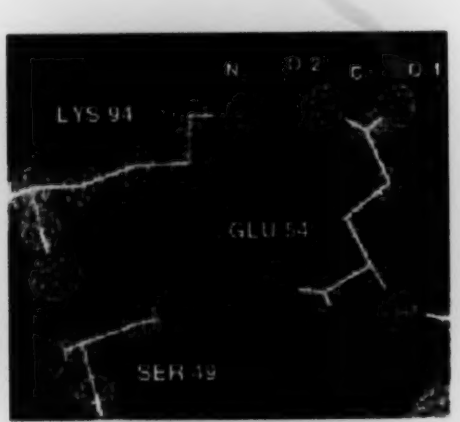

Figure 1. 23 Carbon, nitrogen, and oxygen atoms can be differentiated at ultra-high resolution.

C, N, and O atoms can be differented on the basis of electron density in the serine protease at 0.78 Å. 2Fo-Fc map is contoured at 1 δ - blue and 4 δ -yellow. In lysine 84 the Nζ atom is larger than the carbon atom density. In glutamic acid 53 the oxygen atoms are larger than the carbon atoms. Reprinted with permission from (Kuhn et al., 1998). Copyright (2001) American Chemical Society.

1.4.2.3 Cryocrystallography and Radiation Damage

If room temperature crystals are exposed to the x-ray beam their life-time is often very short, making the collection of an entire data set difficult. This short life-time is the result of x-ray radiation damage to the crystals. Collecting diffraction data from crystals that are cryocooled increases their lifetime substantially; this is the prime purpose of cryocrystallography (Hope, 1990; Rodgers, 1997; Garman and Schneider, 1997). For example, the lifetime of a crystal on a synchrotron source can be increased by cryocooling from a minute to over an hour.

Prior to the mid 1990's data was routinely collected from room temperature crystals. During data collection the crystals were enclosed in a thin-walled glass or quartz capillary along with some mother liquor. Because of radiation damage to the room temperature crystals, multiple crystals would commonly be needed to collect a complete data set. This posed two problems to the crystallographer, firstly merging data collected from different crystals was not always easy or even possible, and secondly crystals were not always easy to produce and were usually limited in number (Hope, 1990).

1.4.2.3.1 History of Cryocrystallography

Macromolecular cryocrystallography's first rudimentary trials began in 1966 with the cryocooling of insulin crystals (Low *et al.*, 1966). From that time it took 30 years for crystal cryocooling to become widely used. The impetus for its increase in popularity came with the introduction of crystal mounting in oil on a glass fiber in the late 1980's (Hope, 1988), and crystal mounting in mother liquor and cryoprotectant in a small loop in 1990 (Teng, 1990). Following these achievements, the percent of structures determined with cryogenic methods increased during the 1990's from less that 1% to more than 40% (Garman and Schneider, 1997; Garman, 2003). Now in 2005 it is expected that close to 90% of all data are collected from cryo-cooled crystals (Garman, 2003). See figure 1.24 below for a survey of cryo versus room temperature data collection for structures published in *Acta Crystallographica D* (Garman, 2003).

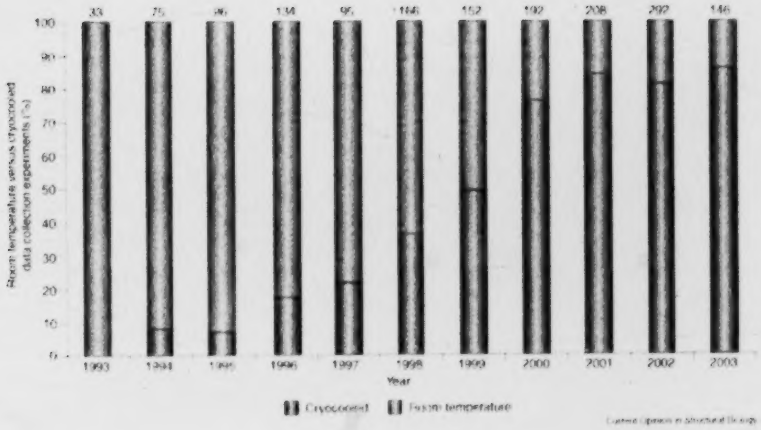

Figure 1. 24 Survey of cryo versus room temperature data collection.

The percentages of room temperature and cryotemperature data collection experiments from a survey of all papers published in *Acta Crystallographica D* from January 1993–May 2003. Reprinted from *Curr Opin Struct Biol*, 13, Garman. 'Cool' crystals: macromolecular cryocrystallography and radiation damage, 545. Copyright (2003), with permission from Elsevier.

The current technique that is most popular for crystal cryocooling involves soaking the crystal briefly in a mother liquor/ cryoprotectant mix to protect it from ice formation, mounting it in a thin fiber loop, then flash-cooling it in gaseous or liquid cryogen such as nitrogen. Subsequently, during data-collection the crystal is maintained at the cryotemperature (Teng, 1990).

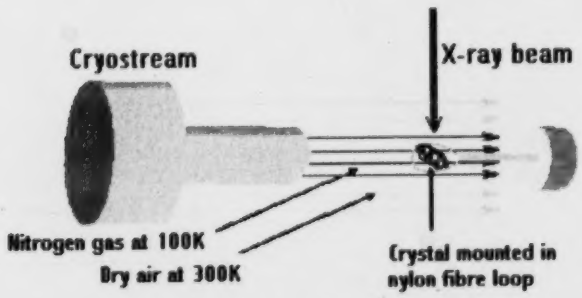

Figure 1.25 Crystal cryocooling

(Rezáèová, 2002).

Not only does crystal cryocooling provide protection from radiation damage, there are additional advantages and a few, usually minor, disadvantages. Advantages include, the molecular and atomic motion (B factor) decrease significantly upon cooling, resulting in a more ordered system, and improving data quality and resolution (Earnest et al., 1991; Garman and Schneider, 1997). The background scatter and absorption is reduced, increasing the signal to noise ratio as a result of the different crystal mounting techniques (Garman and Schneider, 1997). The crystals can be flash cooled and stored in peak condition for long periods of time (Garman, 2003). The disadvantages include, the necessity to establish cryoconditions, and the increase in crystal mosaicity on cryocooling (Garman, 2003).

1.4.2.3.2 Theory of Radiation Damage

To better understand how crystal cryocooling provides protection from radiation damage, the theory of radiation damage will be examined first. There are two main types of radiation damage, primary and secondary. Primary radiation damage is caused by a direct interaction between the radiation beam and the protein (see figure 1.26). A photon from the beam is absorbed by the protein causing the ejection of energetic electrons via the photoelectric, Auger, or Compton effects (Teng and Moffat, 2000). This energy is deposited in small areas of the protein called spurs. The end result of primary radiation damage include the formation of radicals such as hydroxyl radicals, electrons, and hydrogen radicals, as well as the breakage of chemical bonds. This type of radiation damage is independent of crystal temperature and takes place on a picosecond time scale (Teng and Moffat, 2000; Ravelli and McSweeney, 2000; O'Neill *et al.*, 2002). Secondary radiation damage is temperature dependent and takes place on a micro to millisecond time scale. It is caused by radicals generated by the primary damage propogating damage at secondary sites (Teng and Moffat, 2000). Cryocooling affords crystals protection because it prevents diffusion (Garman, 2003). As a result, only one crystal is required to collect a full data set on a rotating anode source, and frequently only one crystal is required at a synchrotron source. This also means that protein crystals with high susceptibility to radiation damage should be ammenable to structural determination.

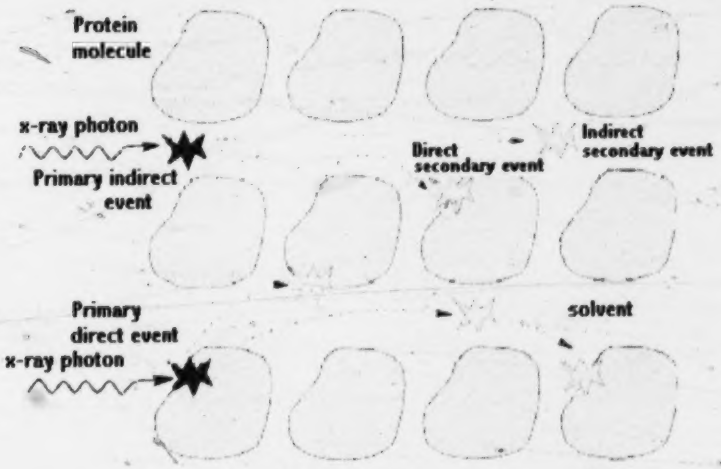

Figure 1.26 Categories of radiation damage to a crystal.

Primary (dark stars), secondary (light stars), direct (on the protein), and indirect (within the solvent) (Murray and Garman, 2002).

Despite the successful reduction in radiation damage that came about as a result of crystal cryocooling, by the mid 1990's with the introduction of 3rd generation synchrotrons radiation damage to cryocooled crystals had again become an issue (figure 1.27) (Garman and Nave, 2002; Teng and Moffat, 2002).

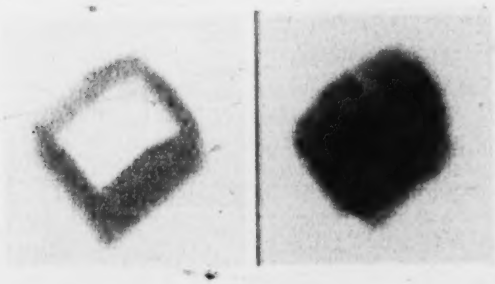

Figure 1.27 Effects of radiation damage on a crystal.

(A) Cryocooled crystal prior to data collection at a 3rd generation synchrotron (B) The same crystal as seen in A, post data collection (Teng and Moffat, 2000).

Aside from the obvious deterioration of the crystal as seen in figure 1.27, before the crystal reaches this stage there are a number of non-specific markers of radiation damage that appear in the diffraction data. This decay in data quality appears initially as a reduction in diffracting power with the high-resolution reflections being the first data to disappear leading to a general increase in atomic B factors (figure 1.28) (Weik *et al.*, 2000; Garman and Nave, 2002; Teng and Moffat, 2002), a decrease in the signal to noise ratio (Teng and Moffat, 2002), and an increase in R_{merge} (Ravelli and McSweeney, 2000). The direct effects on the crystal include an increase in the unit cell volume (figure 1.29) and an increase in mosaicity (Teng and Moffat, 2000; Ravelli and McSweeney, 2000).

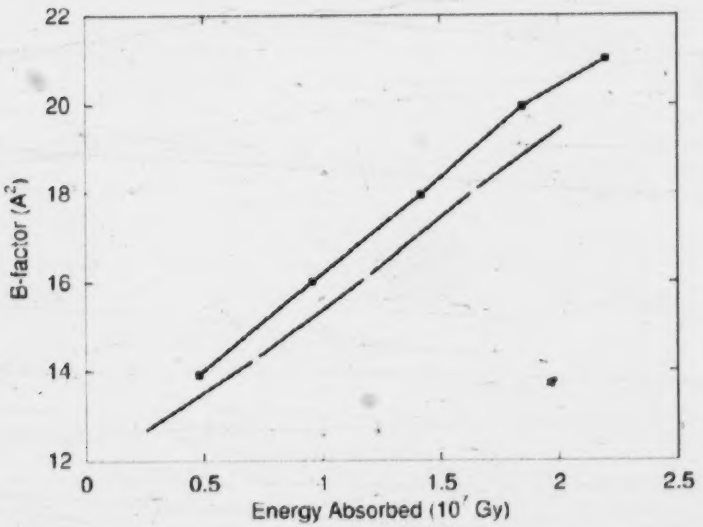

Figure 1.28 Increase in overall B-factors for two cryocooled crystals as a function of absorbed dose (Teng and Moffat, 2002).

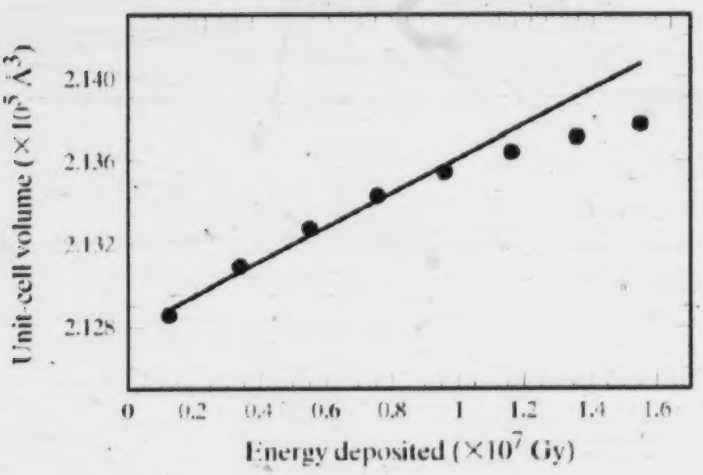

Figure 1.29 Unit-cell volume increase as a function of deposited radiation dose (Teng and Moffat, 2000).

As well as the non-specific effects of radiation damage on crystals, specific residues within the protein may also be affected. These specific effects include disulphide bond cleavage (Burmeister, 2000; Ravelli and McSweeney, 2000; Weik *et al.*, 2000), decarboxylation of acidic residues (Burmeister, 2000; Ravelli and McSweeney, 2000; Weik *et al.*, 2000), alterations of cysteine, methionine, and tyrosine residues (Burmeister, 2000; Carugo and Carugo, 2005), and reduction at metal centers (Carugo and Carugo, 2005). Caution should be used in interpreting structural features at these sites if the crystal has been exposed to a high radiation dose (Garman and Nave, 2002) especially when these residues are involved in crucial functional roles of the protein (Carugo and Carugo, 2005). Chapter four of this thesis shows that the disulphide bond in the PAK pilin was damaged by second and third generation synchrotron radiation (Dunlop and Hazes, 2004). As a result, disulphide bond damage will be discussed in more detail here, followed by a discussion of radiation damage reduction strategies.

The markers of disulphide bond damage include elongation of the disulphide bond and decrease in the electron density for one or both of the Sγ atoms. An example of decreasing electron density over time for the Sγ atoms is shown in figure 1.30. Not all

disulphide bonds have equal susceptibility to radiation damage. Their individual susceptibilities are thought to be influenced by such factors as solvent accessibility, the immediate structural environment, and tertiary structure (Weik *et al.*, 2000).

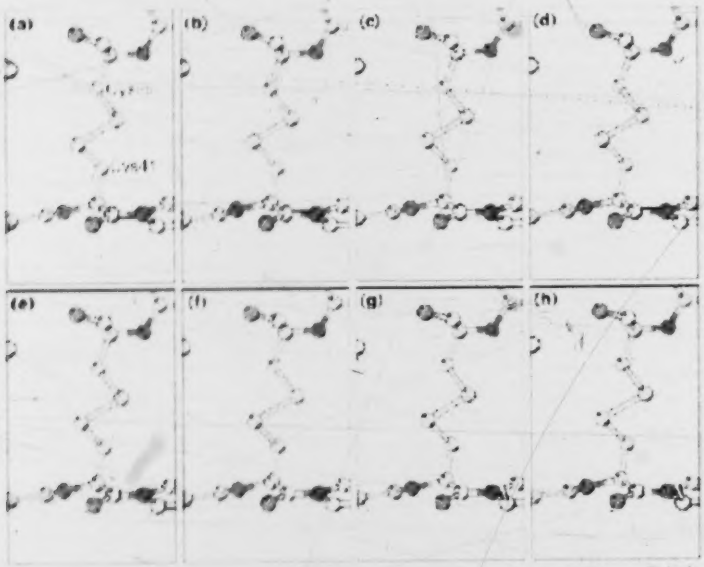

Figure I.30 Difference Fourier maps showing time course cleavage of a disulphide bond

Reprinted from *Structure Fold Des*, 8, Ravelli, McSweeney. The 'fingerprint' that X-rays can leave on structures, 315. Copyright (2000), with permission from Elsevier.

When viewing a number of damaged disulphide bonds in different structures it becomes obvious that they show differing degrees of damage. One aspect of this is the differing Sγ - Sγ bond lengths which can be seen in particular at high resolution. Whereas the normal Sγ - Sγ disulphide bond length is 2.03 Å, upon radiation damage the length may increase indicating that a disulphide radical has formed (Bergés *et al.*, 2000; Weik *et al.*, 2002). There are several disulphide bond radicals that may form. An Sγ - Sγ distance of 2.8 Å indicates that a radical anion RSSR·- has formed following the capture of an electron (Weik 2002; Berges *et al.*, 2000). This radical anion may subsequently be protonated to RSSRH· (Weik 2002) with a bond length of 3.5 Å (Berges *et al.*, 2000). Upon bond

rupture this dissociates into a thiol RSH and a thiyl radical RS· (Weik *et al.*, 2002). Measurement of the Sγ - Sγ bond distance in the PAK pilin structure presented in chapter 4 of this thesis suggests that a disulphide radical has been induced by radiation damage (Dunlop and Hazes, 2005).

1.4.2.3.3 Minimizing Radiation Damage

As a result of the effects of radiation damage at 3rd generation synchrotrons, damage minimization measures are generally taken, however these are not always sufficient. Radiation damage minimization measures include the following. (1) Beam attenuators are used to reduce the intensity of the beam. (2) Data are collected in the most efficient manner possible with minimal radiation dose. Both strategies 1 and 2 are used at the expense of not fully exploiting the full flux of the synchrotron beam (Murray and Garman, 2002). (3) More than one crystal may be used or the crystal may be translated to collect a complete data set. However, merging data from these different sources may be difficult (Murray and Garman, 2002). The ultra high resolution PAK data sets (chapter 4 of thesis) were collected at 2nd and 3rd generation synchrotrons in 1998 before specific radiation damage was widely known to be a problem. In chapter 4, evidence for both specific and non-specific structural radiation damage to the ultra-high resolution PAK structures will be presented (Dunlop and Hazes, 2005).

If the 3rd generation synchrotron beam flux is to be fully exploited to collect the highest resolution data sets possible without causing specific or non-specific radiation damage, the process of radiation damage needs to be fully understood and alternate methods of radiation damage minimization designed (Weik *et al.*, 2002). There are several strategies that are either theoretical or currently under development that have the potential to decrease radiation damage without decreasing the x-ray dose to the crystal. The addition of radical scavengers, in particular ascorbate, to the crystals seems promising. Ascorbate is a radical scavenger that has been shown to decrease radiation damage. Its exact mechanism of action is unknown but it is thought to restore radicals by donating an electron to them (Murray and Garman, 2002; O'Neill *et al.*, 2002). The use of helium

cryostats to lower the crystal temperature to below 100 K has been shown to decrease B-factors, improve data quality, and extend crystal lifetimes (Teng and Moffat, 2002; Murray and Garman, 2002). The use of alternative cryoprotectants could also play a role in reducing radiation damage (Murray and Garman, 2002).

A particularly promising method, named zero-dose extrapolation that rather than minimizing or protecting against radiation damage, computationally corrects intensities that have been damaged by radiation damage, has been developed recently. The damaged intensities are corrected during data reduction. It is hoped that zero-dose extrapolation will be incorporated into data-reduction and scaling programs in the future (Diederichs *et al.*, 2003).

1.4.2.3.4 Structural Changes Resulting From Flash Cooling

In the absence of radiation damage, the structures produced from cryogenic and room temperature data collection are generally assumed to be identical. However, this assumption relies on the idea that flash-cooling traps the protein and solvent in the room temperature conformation; in reality this is not the case. Because characteristic flash-cooling times are in the order of 0.1-1 second and as protein movements are much more rapid, flash cooling should be viewed as a continuous temperature jump relaxation (Halle, 2004). To date, a number of cryocooling-induced structural changes have been documented including decrease in crystallographic B-factors (Earnest *et al.*, 1991), and contraction in unit cell size (Frauenfelder *et al.*, 1987; Parak *et al.*, 1987; Earnest *et al.*, 1991; Tilton *et al.*, 1992) by rearrangement in the packing and orientation of the protein molecules (Juers and Matthews, 2001) and structural changes in the individual protein molecules (Juers and Matthews, 2001), including changes to the conformations of the side chains (Parak et al., 1982 Tilton *et al.*, 1992). In addition changes in the number of visible bound waters have been noted (Carugo and Bordo, 1999; Nakasako 1999). The temperature related changes in the PAK structure are a major subject of study in chapter 4 of this thesis (Dunlop and Hazes, 2005).

1.5 Conclusion to Introduction

Protein crystallography has come a long way since the days when every structure was a time consuming and technically demanding labor of love carried out by highly specialized researchers. In particular, the decade between 1990 and 2000 saw significant progress in critical methods, including third generation synchrotrons, routine crystal cryocooling, MAD phasing, selenomethionine-substituted proteins, and enormous improvements in computer hardware and software. These advances coincided with important improvements in molecular biology including new cloning, expression and purification technologies. As a consequence the field became less specialized and many new researchers entered the field. This all culminated around 2000 with the creation of many high-throughput structural genomics initiatives. High throughput has been applied to every stage of the crystallography process from cloning to refinement of the final structure. In particular, high throughput has made a huge impact on protein crystallization, the stage of crystallography that had become the main bottleneck. Miniaturization of crystallization set ups and automation has enabled crystals to be produced from small quantities of protein and many more crystallization trials to be undertaken and monitored. High throughput methodology not only increased capacity, automation also reduced the need for technical knowledge. Ultimately, it would be desirable for crystallography to become such a routine process that every scientist would have access to it. For this to come to pass the crystallographic process definitely needs to become easier, with every step routinely automated. In addition the strategic planning process for each protein would need to be taken care of by computer programs and all stages branched such that even the difficult cases could be channeled towards success at every stage.

1.6 References

Asherie, N. (2004). *Methods.* **34**, 266-272.

Bard, J., Ercolani, K., Svenson, K., Olland, A., and Somers, W. (2004). *Methods.* **34**, 329-347.

Baruchei, J., Härtwig, J., and Pernot-Rejmánková. (2002). *J. Synchrotron. Rad.* **9**, 107-114.

Berges, J., Fuster, F., Jacquot, J.P., Silvi, B., and Houee-Levin, C. (2000). *Nukleonika*, **45**, 23-30.

Bergfors, T. (2003). *J Struct Biol.* **142**, 66-76.

Bern, M., Goldberg, D., Stevens, and R.C., Kuhn, P. (2004). *J. Appl. Cryst.* **37**, 279-287.

Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535-542.

Bertone, P., Kluger, Y., Lan, N., Zheng, D., Christendat, D., Yee, A., Edwards, A.M., Arrowsmith, C.H., Montelione, G.T., and Gerstein, M. (2001) *Nucleic Acids Res.* **29**, 2884-2898.

Betzel, C., Gourinath, S., Kumar, P., Kaur, P., Perbandt, M., Eschenburg, S., and Singh, T.P. (2001). *Biochemistry.* **40**, 3080-3088.

Blundell, T.L., and Johnson, L.N. "Protein Crystallography". New York: Academic Press, 1976.

Bodenstaff, E.R., Hoedemaeker, F.J., Kuil, M.E., Vrind, H.P.M., and Abrahams, J.P.

(2002). *Acta Cryst* **D58**, 1901-1906.

Burley, S.K., Almo, S.C., Bonanno, J.B., Capel, M., Chance, M.R., Gaasterland, T., Lin, D., Sali, A., Studier, F.W., and Swaminathan, S. (1999). *Nat Genet.* **23**, 151-7.

Burmeister, W.P. (2000). *Acta Cryst.* **D56**, 328-341.

Carter CW Jr, and Carter CW (1979) *J. Biol. Chem.* **254**, 12219-23.

Carugo, O. and Bordo, D. (1999). *Acta Cryst.* **D55**, 479-483.

Chayen, N.E. Shaw Stewart, P.D., and Blow, D.M., (1992) *J. Cryst. Growth* 176–180.

Chayen, N.E., and Saridakis, E. (2002). *Acta Cryst* **D58**, 921-927.

Chayen, N.E. (2003). *J Struct Funct Genomics.* **4**, 115-120.

Chernov, A.A. (2003). *J Struct Biol.* **142**, 3-21.

Collins, F. S., Green, E. D., Guttmacher, A. E. and Guyer, M. S. (2003) *Nature* **422,** 835–847.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760-763.

Dale, G.E., Oefner, C., and D'Arcy, A. (2003). *J Struct Biol.* **142**, 88-97.

D'Arcy, A., Mac Sweeney, A, and Haber, A. (2004). *Methods.* **34**, 323-328.

Declercq, J.P., Evrard, C., Lamzin, V., and Parello, J. (1999). *Protein Sci.* **8**, 2194-2204.

DeLucas, L.J., Hamrick, D., Cosenza, L., Nagy, L., McCombs, D., Bray, T., Chait, A., Stoops, B., Belgovskiy, A., Wilson, W.W., Parham, M., and Chernov, N. (2005). *Prog Biophys Mol Biol.* **88**, 285-309.

Derbin, S.D., and Feher, G. (1996). *Annu. Rev. Phys. Chem.* **47**, 171-204.

Derewenda, Z.S. (2004). *Methods*. **34**, 354-363.

Diederichs, K., McSweeney, S., and Ravelli, R.B.G. (2003). *Acta Cryst*. D**59**, 903-909.

Dunlop, K.V., and Hazes, B. (2003). *Acta Cryst*. D**59**, 1797-1800.

Dunlop, K.V., and Hazes, B. (2005). *Acta Cryst.* D**61**,1041-1048.

Dunlop, K.V., Irvin, R.T., and Hazes, B. (2005). *Acta Cryst.* D**61**, 80-87.

Earnest, T., Fauman, E., Craik, C.S., and Stroud, R. (1991). *Proteins.* **10**, 171-187.

Echalier, A., Glazer, R.L., Fulop, V., and Geday, M.A. (2004). *Acta Cryst* D**60,** 696-702.

Frauenfelder, H. F., Hartmann, H., Karplus, M., Kuntz, I.D. Jr, Kuriyan , J., Para, F., Petsko, G.A., Ringe, D., Tilton, R.F. Jr, Connolly, M.L., and Max, N. (1987). *Biochemistry*. **26**, 254-261.

French, G.S., and Wilson, K.S. (1978). *Acta. Cryst.* A**34**, 517-.

Gao, W., Li, S., and Bi, R. (2005). *Acta Cryst* D**61,** 776-779.

Garcia-Ruiz, J.M. (2003). *J Struct Biol.* **142**, 22-31.

Garman, E. (2003). *Curr Opin Struct Biol.* **13**, 545-551.

Garman, E. and Nave, C. (2002). *J Synchrotron Rad.* **9**, 327-328.

Garman, E.F. and Schneider, T.R. (1997). *J. Appl. Cryst.* **30**, 211-237.

Hajdu, J.(2000). *Curr Opin Struct Biol.* **10**, 569-573.

Hansen, C,. and Quake, S.R. (2003). *Curr Opin Struct Biol.* **13**, 538-544.

Hazes, B., and Price, L. (2005). *Acta Cryst* D**61**, 1165-1171.

Helliwell, J.R. (2005). *Acta Cryst* D**61**, 793-798.

Helliwell, J.R. (1998). *Nat Struct Biol.* **Suppl,** 614-617.

Hendrickson, W.A. *TIBS.* **25**, 637-643.

Hope, H. (1985). *Am Crystallogr. Assoc. Meet. Abstr.* **13,** 24.

Hope, H. (1988). *Acta Cryst.* B**44**, 22-26.

Hope, H. (1990). *Annu. Rev. Biophys. Biophys. Chem.* **19**, 107-126.

Howard, E.I., Sanishvili, R., Cachau, R.E., Mitschler, A., Chevrier, B., Barth, P., Lamour, V., Van Zandt, M., Sibley, E., Bon, C., Moras, D., Schneider, T.R., Joachimiak, A., and Podjarny, A. (2004). *Proteins.* **55**(4), 792-804.

Hui, R., and Edwards, A. (2003). *J Struct Biol.* **142**, 154-61.

Jacquamet, L., Ohana, J., Joly, J., Borel, F., Pirocchi, M., Charrault, P., Bertoni, A., Israel-Gouy, P., Carpentie, P., Kozielski, F., Blot, D., and Ferrer, J.L. (2004). *Structure* **12**, 1219-1225.

Jancarik, J. and Kim, S.H. (1991). *J.Appl.Crystallogr.* **24**, 409.

Jelsch, C., Teeter, M.M., Lamzin, V., Pichon-Pesme, V., Blessing, R.H., Lecomte, C. (2000). *Proc Natl Acad Sci U S A.* **97**(7), 3171-3176.

Judge, R.A., Snell, E.H., and van der Woerd, M.J. (2005). *Acta Cryst.* **D61**, 763-771.

Kadri, A., Lorber, B., Charron, C., Robert, M.C., Capelle, B., Damak, M., Jenner, G., and Giege, R. (2005). *Acta Cryst.* **D61**, 784-788.

Kierzek, A.M., and Zielenkiewicz, P. (2001). *Biophys Chem.* **91**, 1-20.

Krupka, H.I., Rupp, B., Segelke, B.W., Lekin, T.P., Wright, D., Wu, H.C., Todd, P., and Azarani.A. (2002). *Acta Cryst* **D58,** 1523-1526.

Kuhn, P., Knapp, M., Soltis, M., Ganshaw, G., Thoene, M., and Bott, R. (1998). *Biochemistry.* **37**, 13446-13452.

Lecomte, C., Guillot, B., Muzet, N., Pichon-Pesme, V., and Jelsch, C. (2004). *Cell. Mol. Life. Sci.* **61**, 774-782.

Leulliot, N., Tresaugues, L., Bremang, M., Sorel, I., Ulryck, N., Graille, M., Aboulfath, I., Poupon, A., Liger, D., Quevillon-Cheruel, S., Janin, J., and van Tilbeurgh, H. (2005) *Acta Cryst.* **D61**, 664-670.

Lin,T., Schildkamp, W., Brister, K., Doerschuk, P.C., Somayazulu, M., Mao, H.K., and Johnson, J.E. 2005. *Acta Cryst.* **D61**, 737-743.

Luft, J.R., Collins, R.J., Fehrman, N.A., Lauricella, A.M., Veatch, C.K., and Detitta, G.T. (2003). *J Struct Biol.* **142**, 170-179.

MacArthur, M.W., and Thornton, J.M. (1999). *Acta Cryst.* **D55**, 994-1004.

Mattos, C. (2002). *TIBS.* **27**, 203-208.

McPherson, A. (2004a). *J Struct Funct Genomics.* **5**, 3-12.

McPherson, A. (2004b). *Methods.* **34**, 254-265.

McPherson, A. (1999).*Crystallization of Biological Macromolecules,* Cold Spring Harbor Laboratory Press, New York.

McPherson, A., Kuznetsov, Y.S., and Sayre, D. (2003). *J Struct Biol.* **142**, 32-46.

McRee, D.E. (1999). *J. Struct. Biol.* **125**, 156-165.

Miao, J., Hodgson, K.O., and Sayre, D. (2001). *PNAS.* **98**, 6641-6645.

Murray, J., and Garman, E. (2002). *J. Synchrotron Rad.* **9**, 347-354.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). *Acta Cryst.* **D53**, 240-255.

Nakasako, M. (1999). *J. Mol. Biol.* **289**, 547-564.

Nakasako, M. (2001). *Cellular and Molecular Biology.* **47 (5)**, 767-790.

Nanev, C.N., Penkova, A., and Chayen, N. (2004). *Ann. N.Y. Acad. Sci.* **1027**, 1-9.

O'Neill, P., Stevens, D.L., and Garman, E.F. (2002). *J. Synchrotron Rad.* **9**, 329-332.

Page, R and Stevens, R.C. (2004). *Methods*. **34**, 373-389.

Parak, E.N., Frolov, R.L., Mossbauer., and Goldanskii, V. I. (1981). *J. Mol. Biol.* **145**, 825-833.

Parak, E.N., Hartmann, H., Aumann, K.D., Reuscher, H., Rennekamp, G., Bartunik, H., and Steigemann, W. (1987). *Eur Biophys J.* **15**, 237-249.

Pusey, M.L., Liu, Z.J., Tempel, W., Praissman, J., Lin, D., Wang, B.C., Gavira, J.A., and Ng, J.D. (2005). *Prog. Biophys. Mol. Biol.* **88**, 359-386.

Ravelli, R.B. and McSweeney, S.M. (2000) *Struct. Fold. Des.* **8**, 315-328.

Rayment, I. (1997) *Methods Enzymol.* **276,** 171-179.

Rezáèová, P. (2002) *Materials Structure.* **9,** 15-17.

Ringe, D., and Petsko, G.A. (2003). *Biophys. Chem.* **105**, 667-680.

Rodgers, D.W. (1997). *Methods Enzymol.* **276,** 183-203.

Rupp, B. (2003). *J. Struct. Biol.* **142**, 162-169.

Rupp, B., Wang, J. (2004). *Methods*. **34**, 390-407.

Santarsiero, B.D., Yegian, D.T., Lee, C.C., Spraggon, G., Gu, J., Scheibe, D., Uber, D.C., Cornell, E.W., Nordmeyer, R.A., Kolbe, W.F., Jin, J., Jones, A.L., Jaklevic, J.M., Schultz, P.G., and Stevens, R.C. (2002). *J. Appl. Cryst.* **35**, 278-281.

Scott, W. G., Finch, J.T., Grenfell, R., Fogg, J., Smith, T., Gait, M.J., and Klug, A.,

(1995) *J Mol Biol.* **250**, 327-332.

Segelke, B.W. (2000). *ACA Meeting Series.* **27**, 43.

Segelke, B.W. (2001). *J Cryst Growth.* **232**, 553-562.

Segelke, B.W., Schafer, J., Coleman, M.A., Lekin, T.P., Toppani, D., Skowronek, K.J., Kantardjieff, K.A., and Rupp, B. (2004) *J Struct Funct Genomics.* **5**, 147-57.

Spirin, A.S. (2004). *TRENDS in Biotechnology.* **22**, 538-545.

Spraggon, G., Lesley, S.A., Kreusch, A., and Priestle, J.P. (2002). *Acta Cryst* D**58**, 1915 - 1923.

Stryer, L. (1995). *Biochemistry.* 4th edition, WH Freeman and Company.

Stura, E.A., Nemerow, G.R., and Wilson, I.A. (1992). *J Cryst Growth.* **122**, 273-285.

Taylor, G. (2003). *Acta. Cryst.* D**59**, 1991-1890.

Teng, T.Y. (1990). *J. Appl. Cryst.* **23**, 387-391.

Teng, T., and Moffat, M. (2000) *J. Synchrotron Rad.* **7**, 313-317.

Teng, T., and Moffat, M. (2002) *J. Synchrotron Rad.* **9**, 198-201.

Tilton, R.F., Dewan, J.C., and Petsko, G.A. (1992). **31**, 2469-2481.

Vaguine, A. A., Richelle, J. and Wodak, S. J. (1999). *Acta Cryst.* **D55,** 191-220.

Vergara, A., Lorber, B., Zagari, A., and Giege, R. (2003). *Acta Cryst* D**59**, 2-15.

Vincentelli, R., Bignon, C., Gruez, A., Canaan, S., Sulzenbacher, G., Tegoni, M., Campanacci, V., and Cambillau, C. (2003). *Acc. Chem. Res.* **36**, 165-172.

Walsh, M.A., Schneider, T.R., Sieker, L.C., Dauter, Z., Samzin, V.S., and Wilson, K.S. (1998). *Acta. Cryst.* **D54,** 522-546.

Walter, T.S., Diprose, J.M., Mayo, C.J., Siebold, C., Pickford, M.G., Carter, L., Sutton, G.C., Berrow, N.S., Brown, J., Berry, I.M., Stewart-Jones, G.B., Grimes, J.M., Stammers, D.K., Esnouf, R.M., Jones, E.Y., Owens, R.J., Stuart, D.I., and Harlos, K. (2005). *Acta Cryst.* **D61**, 651-657.

Weik, M., Ravelli, R.B., Kryger, G., McSweeney, S., Raves, M.L., Harel, M., Gros, P., Silmna, I., Droon, J., and Sussman, J.L. (2000). *Proc. Natl Acad. Sci. USA*. **97**, 623-628.

Weik, M., Berges, J., Raves, M.L., Gros, P., McSweeney, S., Silman, I., Sussman, J.L., Houee-Levin, C., and Ravelli, R.B.G. (2002). *J. Synchrotron Rad.* **9**, 342-346.

Wilson, J. (2002). *Acta Cryst* **D58**, 1907-1914.

Wilson, W.W. (2003). *J Struct Biol.* **142**, 56-65.

Yee, A., Pardee, K., Christendat, D., Savchenko, A., Edwards, A.M., and Arrowsmith, C.H. (2003). *Acc. Chem. Res.* **36**, 183-189.

When less is more: a more efficient vapour-diffusion protocol

Published in *Acta Crystallogr D Biol Crystallogr.* 2003 Oct;59 (Pt 10):1797-800.

Authors: Dunlop, K.V and Hazes B.

Reproduced with kind permission from the International Union of Crystallography.

Chapter 2: When Less Is More: A More Efficient Vapour Diffusion Protocol

2.1 Introduction

2.1.1 Reducing Protein Consumption

Reducing protein consumption during crystallization screening is an important challenge for crystallographers (Rupp, 2003a). This is because the availability of sufficient amounts of highly purified protein, in particular for those proteins that are difficult to overexpress, is frequently a limiting factor. For traditional vapour diffusion crystallization screening, the desired amount of protein before screening is started is usually 5-10mg, although often much greater amounts are used before a diffraction quality crystal is obtained.

There are three factors that determine the total amount of protein used in a crystallization screen. These factors are, the number of crystallization trials, the volume of protein used per trial, and the concentration of protein used per trial. This is illustrated in the equation below.

total protein = no. of trials x volume/trial x concentration/trial

As a consequence, there are three possible means to reduce protein consumption during crystallization screening. One involves reducing the number of crystallization trials by more efficiently sampling crystallization space. There are a couple of means to achieve this. a) Using a random sampling screening technique; this has been found to be the most efficient screening method (Segelke, 2001). b) Doing a small pre-screen to determine the protein solubility at a range of pH's.This is useful so that an optimal choice of screening conditions may be made (Rupp, 2003b). In future, the results from data mining operations that correlate properties of proteins and their crystallization conditions may allow the

rational prediction of crystallization conditions for previously uncrystallized proteins (Luft *et al.*, 2003) thus reducing the number of crystallization trials required.

The second method to reduce protein consumption is decreasing the volume of protein solution used per trial. As the limitation of manual pipetting is approximately 0.5µL, robots are required for dispensing smaller volumes. Santarsiero *et al.*, (2002) has reported producing diffraction-quality crystals using as little as 40nL hanging drop size. However, 100nL volumes more routinely produce diffraction quality crystals (Santarsiero *et al.*, 2002). This still represents a ten-fold improvement over manual pipetting. Unfortunately, the robots are very expensive, and a patent prevents setting up crystallization drops with volumes less than 1µL (Regents of the University of California, 2001).

The third method of reducing protein consumption involves dilution of the protein solution. Here we demonstrate a modified vapour diffusion method in which the protein and mother liquor in the crystallization drop are diluted.

2.1.2 The Dilution Method

The theoretical differences between the traditional vapour diffusion method and the dilution method are illustrated in figure 2.1. In a traditional experiment, the initial concentration of protein ($^1/_2C_P$) and mother liquor ($^1/_2C_{ML}$) (black dot) falls into the undersaturated region of crystallization space. The notation, $^1/_2C$, describes the halving of the initial concentrations of protein and mother liquor upon formation of the crystallization drop. By vapour diffusion both the protein concentration and mother liquor concentration increase (bold black line) until the concentrations pass into the nucleation region where a nucleation event can provide the core for crystal formation. During the equilibration process the concentration of the mother liquor in the drop doubles. The protein concentration also rises until the nucleation event at which point the protein concentration in solution decreases as the crystal is formed.

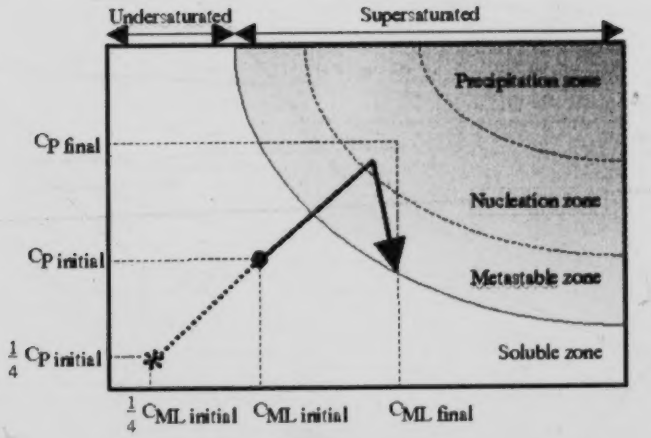

Figure 2. 1 Crystallization space diagram showing the path of mother liquor and protein concentrations during a vapour diffusion crystallization experiment.

The black circle at $C_{ML\ initial}$ and $C_{P\ initial}$ represents the initial mother liquor and protein concentrations for traditional vapour diffusion crystallization experiments. The asterisks at $1/4C_{ML\ initial}$ and $1/4C_{P\ initial}$ represents the initial concentrations of protein and mother liquor for a 4-fold diluted vapour diffusion experiment. The dashed and solid lines indicate the concentrations during the equilibration process.

In the dilution method both the protein and precipitant solution are diluted n-fold. The initial conditions for a 4-fold diluted experiment are given by $1/4C_{ML\ initial}$ and $1/4C_{P\ initial}$ in figure 2.1. Due to vapour diffusion the protein and precipitant concentration will increase as indicated by the dashed line. When the concentrations have reached $C_{ML\ initial}$ and $C_{P\ initial}$ respectively, a situation identical to the traditional experiment is obtained but with an n-fold lower drop volume, 4-fold lower in the example in figure 2.1. Here we explore the practical use of this new strategy.

2.2 Experimental

2.2.1 Protein Preparation

Lysozyme (hen egg white, Sigma) (14.3kDa) was dialyzed against 0.1M sodium acetate, pH 4.8 whereas glucose isomerase (Hampton Research) (173kDa) was dialyzed against

water. The final concentrations determined by uv/vis spectroscopy (ε_{280} values 1.042 and 2.68) were 50mg/mL and 26mg/mL, respectively.

2.2.2 Crystal Growth

The crystals were grown in quadruplicate at 295K by sitting drop vapour diffusion using previously determined crystallization conditions. For lysozyme, the reservoir solution consisted of 1mL of 8% w/v sodium chloride, and 0.1M of sodium acetate at a pH of 4.8 (modified from Drenth, 1994). For glucose isomerase, the reservoir solution consisted of 1mL of 0.7M sodium citrate, pH 6.3 (Hampton Research).

To form the crystallization drop 1μL of the diluted protein solution and 1μL of the diluted mother liquor solution were pipetted into the sitting drop post. 1 mL of undiluted reservoir solution was pipetted into the reservoir.

2.3 Results/Discussion

2.3.1 Effects of the Vapour Diffusion Method on Crystallization

As explained in the introduction, crystallization theory predicts that the dilution method is equivalent to ultra-small volume crystallization experiments after the evaporation of the excess water from the drop. In a preliminary test of the method we created crystallization experiments for lysozyme and glucose isomerase with dilutions up to 6-fold. Tables 2.1 and 2.2 give the results of these experiments.

Table 2.1 The effect of the dilution method on lysozyme crystal growth.

Dilution	*Number of crystals/drop*[1]	*Longest crystal edge (μm)*[1]
1	1150 (300-2000)	65 (30- 100)
2	70 (60-80)	125 (110-140)
3	25	120 (110-130)
4	15 (10-20)	140
5	20 (15-25)	120 (110–130)
6	15 (10-20)	130 (110-150)

[1] The mean value for the quadruplicate experiment is listed first followed by the lowest and highest values in parentheses. For lysozyme, all of the quadruplicate experiments yielded crystals.

For lysozyme, crystals were obtained at all dilution levels. The crystals all had the same morphology, however, dilution had an effect on two crystal properties, size and number. The size of the crystals increased almost two-fold upon a two-fold dilution. Further dilution had no effect on the size of the crystals. The number of crystals formed decreased upon increasing dilution between one and three-fold, however, at dilutions greater than three-fold the number remained constant (Figure 2.2; a-f).

Table 2.2 The effect of the dilution method on glucose isomerase crystal growth.

Dilution	*Crystal Form* [1]	*Number of Wells*[2]	*Number of crystals/drop*[3]	*Longest crystal edge (μm)*[3]
1	A	4	16 (10-32)	202 (110-250)
2	A	1	1	210
	B	2	4 (1-7)	700 (600-800)
3	B	4	7 (1-10)	940 (800-1000)
4	B/C	4	10	300 (200-400)
5	B	1	20	1000
	B/C	3	9	470 (400-500)
6	B	1	5	1000
	B/C	2	9 (8-10)	650 (600-700)

[1] A, B, and C represents the 3 different crystal morphologies of glucose isomerase shown in figure 2.2g, 2.2h, and 2.2i, respectively.
[2] The number of wells in the quadruplicate experiment that yielded this crystal form.
[3] The mean value is listed first followed by the lowest and highest values in parentheses.

For glucose isomerase, crystals were also obtained at all dilution levels, but dilution caused the appearance of two new crystal forms. The crystal form depends on the extent of dilution. Form A (Figure 2.2g) has a spacegroup of I222, and crystallized at one and two-fold dilutions. In drops containing crystal form A, no other crystal form was present. Form B (Figure 2.2h) has a spacegroup of $P2_12_12$ and crystallized in drops at dilutions of two-fold or greater. Structures for both Form A and B crystals have been solved previously. Form C (Figure 2.2i) crystals were too small to be characterized, and crystallized in drops at dilutions of four-fold or greater. Crystal forms B and C could appear as a single form in a crystallization drop or as a mix of these two forms. The form B crystals were larger when they were not mixed with form C.

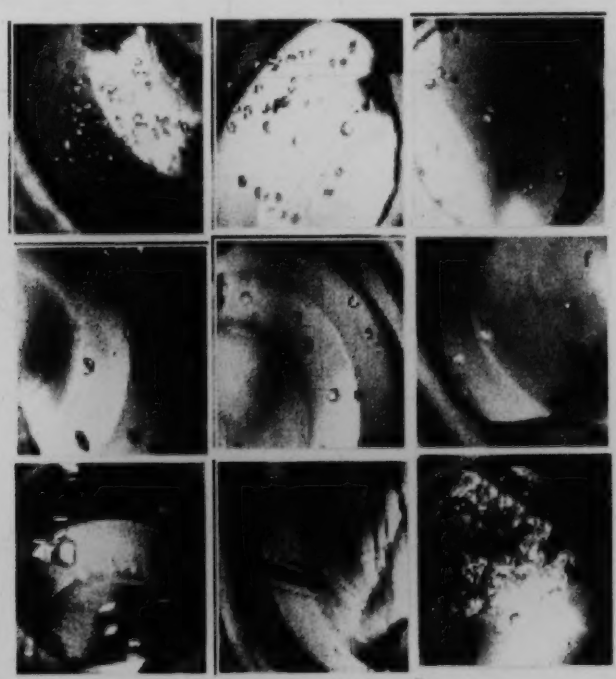

Figure 2.2 Crystals of lysozyme and glucose isomerase.

(a- f) lysozyme crystals at 1 to 6 fold dilutions respectively, (g) glucose isomerase crystal form A, (h) glucose isomerase crystal form B, (i) glucose isomerase crystal form C.

The time taken for crystal growth in the dilution method was slightly longer than for the

traditional method. The extra time taken depended on the dilution of the crystallization drop, but was never more than one day. Considering the usual duration of tracking the progress of a crystallization experiment the increase in time is insignificant.

2.3.2 Dilution and Traditional Vapour Diffusion Method Comparison

As predicted by theory, the dilution method in comparison with the traditional method does not affect crystal detection in our two test cases. There are, however, clear differences between the diluted and undiluted experiments most likely caused by processes occurring during the evaporation of the excess water from the drop (corresponding to the dashed line in Figure 2.1). For instance, if there are volatile components other than water in the mother liquor then they may diffuse to the crystallization droplet and affect crystallization behaviour. More interestingly, if the starting conditions of the undiluted experiments already fall in the nucleation zone, then the diluted experiment will traverse a region of lower supersaturation not explored by the undiluted experiment. Fewer nucleation sites and therefore larger crystals could be expected in this situation. This is indeed what we observe for lysozyme in the undiluted to three-fold diluted range. This suggests that at least the undiluted and two-fold diluted starting conditions are already in the nucleation zone. For glucose isomerase dilution actually leads to different crystal morphologies. Again, this could occur if the initial conditions are already in the nucleation zone and the different crystal forms need different levels of superaturation; crystal form A would need the highest supersaturation and crystal form C the lowest.

2.3.3 Advantages and Concerns of the Dilution Method

Although more research is required, our preliminary results suggest that the dilution method is a practical alternative for ultra-small volume experiments. Apart from the fact that it is amenable to both manual and automated procedures there are a few other practical advantages over ultra-small volume drops. Firstly, small nanoliter volumes can evaporate during the time between dispensing and sealing the plate. Current solutions

involve base plate cooling and humidifying the robotic set-up environment (Bodenstaff *et al.*, 2002). The dilution technique will greatly reduce or eliminate this problem because of the very low initial precipitant concentrations of the drops and a large volume to surface area ratio. Secondly, dispensing accuracy decreases as dispensing volumes decrease, especially for viscous or low surface tension solutions (Bodenstaff *et al.*, 2002). The dilution method improves accuracy by both increasing dispensing volumes and decreasing viscosity. Thirdly, dilution ensures that almost all experiments will start in the undersaturated zone. As a result, there is an increased likelihood of crystal formation. Fourthly, since the dilution method can give different crystallization behaviour, it may be considered as an optimization method when preliminary conditions have already been found. Based on theory and the observations for lysozyme, dilution may lead to fewer and larger crystals. Alternatively, the dilution method may yield different crystal forms that may be more suitable for diffraction studies or can facilitate structure solution by cross-crystal averaging.

One concern when using the dilution method is that dilution may cause protein instability if salt is required for its stability. There was no such problem in our experiments, but the two test proteins are highly soluble. Tests on a broader range of proteins will be undertaken to determine whether this potential concern occurs in practise.

2.4 Conclusions

A simple but innovative vapour diffusion protocol has been developed to reduce protein consumption in crystal screens. The method can be used in manual or automated modes and does not require special equipment. The primary focus of the method is to conserve protein but our theoretical considerations and our observations suggest that it may also have some practical advantages over ultra-small volume experiments. Crystallization experiments on two test proteins have confirmed and refined our theoretical expectations and constitute a proof of concept. A much more extensive study is planned to evaluate the general applicability and to make a more quantitative comparison with existing methods.

2.5 References

Bodenstaff, E.R., Hoedemaeker, F.J., Kuil, M.E., de Vrind, H.P.M., Abrahams, J.P. (2002). *Acta Cryst.* D**58**, 1901-1906.

Drenth, J. *Principles of Protein X-ray Crystallography.* Springer-Verlag New York, Inc, 1994.

Hampton. (2001). *Glucose Isomerase Users Guide.* Hampton Research, Laguna Niguel.

Luft JR, C. R., Fehrman NA, Lauricella AM, Veatch CK, DeTitta GT (2003). *J Struct Biol.* **142**(1): 170-179.

Rupp, B. (2003a). *J Struct Biol.* **142**(1): 162-169.

Rupp, B. (2003b). *Acc Chem Res.* **36**: 173-181.

Santarsiero, B. D., Yegian, D. T., Lee, C.C., Spraggon, G., Gu, J., Scheibe, D., Uber, D.C., Cornell, E.W., Nordmeyer, R.A., Kolbe, W.F., Jin, J., Jones, A.L., Jaklevic J.M., Schultz, P.G., Stevens, R.C. (2002) *J. Appl. Cryst.* **35**: 278-281.

Segelke, B.W. (2001)) *J. Cryst. Growth.* **232**: 553-562.

The Regents of the University of California. (2001). *Methods and apparatus for performing array microcrystallizations.* US Patent, 6 296 673.

A modified vapor-diffusion crystallization protocol that uses a common dehydrating agent

Published in *Acta Crystallogr D Biol Crystallogr*. 2005 Aug;61(Pt 8):1041-8.

Authors: Dunlop K.V and Hazes, B.

Reproduced with kind permission from the International Union of Crystallography.

Chapter 3: A Modified Vapor Diffusion Crystallization Protocol That Uses A Common Dehydrating Agent

3.1 Introduction

The realization that protein structural knowledge is necessary to understand the molecular mechanisms of protein function has led to rapid expansion of the field of protein crystallography. In addition, there have been large methodological improvements in all aspects of structure determination (Abola *et al.*, 2000), with structural proteomics consortia and the pharmaceutical industry playing a particularly important role in driving the developments. Unfortunately, protein crystallization has largely remained a trial-and-error process and growing well-diffracting protein crystals has become a significant bottleneck. The main response has been to increase the number of crystallization trials, exploring more chemical and physical crystallization parameters and targeting multiple related proteins, such as orthologs and mutants (Hui and Edwards, 2003; Mateja *et al.*, 2002). This high-throughput approach has been made possible by the introduction of crystallization robots and automated imaging systems (Hui and Edwards, 2003; Krupka *et al.*, 2002; Spraggon, 2002). To maximize the advantages and minimize the limitations of the new technologies it is important to re-evaluate our crystallization protocols. We recently proposed the *Dilution Method* to improve crystallization results, reduce protein consumption, and minimize robotic dispensing problems caused by viscosity and evaporation issues (Dunlop and Hazes, 2003). Here we revisit the role of the reservoir solution, and show that significant practical benefits can be realized when many different crystallization experiments share a common reservoir solution.

The vapor diffusion crystallization technique is the most widely used protein crystallization method. Its goal is to bring about supersaturation of a protein solution in a physicochemical environment that allows well-ordered crystals to form (McPherson, 1976). Apart from the conditions at the endpoint of the experiment, the kinetics of change

during the vapor diffusion experiment also affect the outcome, resulting in a complex multi-dimensional parameter space of crystallization (McPherson, 2004). Different chemical environments are explored by mixing the protein solution with different crystallization solutions, typically in a 1:1 ratio. The crystallization solution normally contains a precipitant, buffer and one or more additives. Each resulting crystallization drop is then equilibrated in a sealed chamber against a larger volume of reservoir solution with a higher precipitant concentration than the crystallization drop. (We use the terms crystallization solution and reservoir solution to refer to the solutions added to the protein sample and reservoir, respectively. We avoid the term "mother liquor" because it is used indiscriminately for both solutions). As a result, water vapor will diffuse from the drop to the reservoir solution leading to a slow concentrating effect on the crystallization drop. The vapor diffusion process continues until the drop and reservoir solution have the same equilibrium vapor pressure, which equates to them having the same water potential. In cases where water is the only volatile component, any solution with the appropriate water potential can be used as the reservoir solution. However, historically the reservoir solution has been the same as the crystallization solution. Before the introduction of pre-made crystallization screens this was convenient, because the reservoir well of the crystallization plate could be used to mix precipitant, buffer, and additives prior to addition to the crystallization drop. In addition, this procedure ensures that the water activity of the reservoir solution is lower than that of the drop and, therefore, that the drop concentrates during equilibration. Finally, in this approach non-aqueous volatiles are present in both drop and reservoir and, upon equilibrium, the composition in the drop will closely match that of the reservoir.

The current wide availability of pre-made crystallization screens makes it feasible to add the complex screen solutions just to the crystallization drops and use a more convenient solution of the right dehydrating strength to fill the reservoir. This is not a completely new idea; published references include McPherson, who used it to pre-screen proteins for favored pH and precipitant type (McPherson, 1992), Berger, who developed a nucleic acid crystallization screen that equilibrates 24-conditions against a single reservoir solution (Berger *et al.*, 1996; commercialized as the Nucleic Acid Mini Screen by

Hampton Research), and Douglas Instruments, who make the 'Douglas Vapor Batch Plate' that could be used for this method (Mortuza *et al.*, 2004). However, replacing the traditional reservoir solution by an alternate reservoir solution of equivalent dehydrating strength affects crystallization in many ways. Most importantly, the assumption that water is the only volatile component is rarely correct. In addition to obvious volatile components such as isopropanol, other reagents like MPD, ethylene glycol and even salts are volatile to various extents. Since the shared reservoir solution does not contain any of these reagents, they will be transferred from the drop to the reservoir. Similarly, a different pH between drop and reservoir can drive the transfer of acids or bases, especially when more volatile compounds such as acetate, bicarbonate, and ammonia are present. Indeed, transfer of ammonia has been used to induce crystallization (Mikol *et al.*, 1989) and transfer of isopropanol from the reservoir to the crystallization drop was recently used to crystallize a viral capsid protein (Mortuza *et al.*, 2004). The impact of all these effects on the experiment will depend on the rate of transfer relative to the duration of the crystallization experiment. Changing the reservoir solution can also affect equilibrium kinetics (Forsythe *et al.*, 2002). Although this effect may be dominated by evaporation at the drop-air interface, some effect on water absorption at the reservoir-air interface cannot be excluded.

For practical reasons, we are particularly interested in protocols where all experiments share the same reservoir solution, because this allows the design of novel crystallization plates where many drops are equilibrated against one reservoir solution in large open crystallization chambers (Figure 3.1). We have named this modified vapor diffusion crystallization approach the shared reservoir solution (SRS) method. However, because each crystallization solution has a different water activity, the water activity of a shared reservoir solution is no longer directly linked to that of the drop and drops will therefore concentrate, or even dilute, to different extents. Adoption of the new plate format will also change some basic parameters of the experiment. Although the plates can be designed so that the volume of reservoir solution, drops, and air space are proportionally the same as in traditional plates, the distance of drops to the reservoir will be longer, which can affect equilibration kinetics (Fowlis *et al.*, 1988). There is also potential for

chemical cross-talk between drops, and the presence of a large open air space may itself have subtle effects.

Ultimately, the usefulness of the SRS method depends on the success rate of finding protein crystallization conditions, combined with the practical benefits for robotic or manual screening. This can only be determined experimentally. One very recent study, evaluating the method on lysozyme, was accepted for publication while we completed our manuscript (Newman, 2005). That study, which used conventional crystallization plates, indicates that the SRS method actually performs significantly better than the traditional protocol. Here we present a broader test of the method using our novel plates as well as conventional plates. Our results complement and extend the observations by Newman and further indicate that the success rates with our new plates and the conventional plates is equivalent. Furthermore, we discuss the practical consequences for high-throughput protein crystallization in both experiment setup and visualization.

3.2 Experimental

3.2.1 Protein Preparation

Lysozyme (hen egg-white, Sigma; MW=14.3 kDa) was dialyzed against 0.1 M sodium acetate pH 4.8, whereas glucose isomerase (Hampton Research; MW=173 kDa) and thaumatin (from *Thaumatoccus daniellii*, Sigma, MW=22.2 kDa) were dialyzed against water. UV-Vis spectroscopy gave final concentrations for lysozyme, glucose isomerase, and thaumatin of 50, 26, and 50 mg ml^{-1}, respectively (using A280 values of 1.04, 2.68, and 1.25, respectively).

3.2.2 Crystallization

All crystallization experiments were set up at 295 K using sitting-drop vapor diffusion in either CrystalQuick conical bottom 96 well plates or using SRS prototype plates (see

figure 3.1). The Greiner plates have three sitting drop wells per reservoir, which were used to set up the experiment in triplicate. The experimental solutions were dispensed with a Honeybee crystallization robot (Genomic Solutions) using 50 µl of reservoir solution and crystallization drops made of 100 nl of protein sample and 100 nl of crystallization solution. The results of the crystallization experiments were reported based on experiments at day five.

To demonstrate that the SRS method can reproduce crystallization results obtained with traditional methods, we used both methods to crystallize the three test proteins using known crystallization conditions. For lysozyme, the crystallization solution consisted of 8% (*w/v*) sodium chloride and 0.1 M sodium acetate at pH 4.8 (modified from Drenth, 1994). For thaumatin a crystallization solution with 32% sodium tartrate, 5% ethylene glycol, and 0.1 M Hepes pH 7.2 was used (modified from Rigaku MSC, http://www.rigakumsc.com/protein/crystallization.html). For glucose isomerase two crystallization conditions were used: 1) 25% MPD, 0.2 M magnesium chloride, 0.1 M Tris pH 7, and 2) 0.8 M sodium citrate, pH 6.2.

To evaluate the performance of the SRS method in typical crystal screens, we compared it against the traditional method using the PEG ion (Hampton Research), Cryo I (Emerald BioStructures), and a locally developed salt/pH grid screen. The grid screen explores six salts, each at 4 different concentrations (sodium citrate 0.3 - 1.35 M, sodium malonate 1.0 - 3.1 M, sodium bromide 1.5 - 3.6 M, magnesium chloride 0.5 -1.71 M, calcium chloride 0.8 - 2.61 M, and lithium chloride 1 - 7 M) and four buffers (citric acid pH 4.29, MES pH 5.5, Tris pH 6.8, Tris pH 8.3) to make a total of 96 different conditions.

3.3 Results

3.3.1 Choice of Precipitant and Concentration to Use as Shared Reservoir Solution

We hypothesized that any precipitant could be used as SRS, as long as it is of the right

dehydrating strength and water is the only volatile component. To test this hypothesis we compared traditional experiments using known crystallization conditions, with the SRS method. The SRS trials included the known crystallization solution in the crystallization drop and one of three common precipitants as the SRS, viz. $(NH_4)_2SO_4$, NaCl, and PEG 3350. To determine how sensitive the method is to the precipitant concentration of the reservoir solution, a broad range of concentrations were used. The experiments that used $(NH_4)_2SO_4$ as the SRS did not yield any crystals. The results for the other two SRS choices are given in Table 3.1.

Table 3.1 Evaluation of NaCl and PEG 3350 as Shared Reservoir Solutions

Precipitant	Lysozyme	Glucose isomerase 1	Glucose isomerase 2	Thaumatin
Control†	20/132/3	1/170/1	3/50/3	3/60/2
0.5 *M* NaCl	0	2/99/2	2/90/2	3/40/2
1 *M* NaCl	30/120/2	2/125/1	1/80/3	12/180/3
1.5 *M* NaCl	30/120/3	0	2/70/3	10/350/3
2 *M* NaCl	1/519 cluster/2	0	2/50/2	200/120/3
3 *M* NaCl	0	0	3/50/2	0
4 *M* NaCl	0	0	4/40/2	10/20/1
5.00%(w/v) PEG 3350	0	2/60/2	1/100/3	20/5/1
10.00%(w/v) PEG 3350	0	0	1/100/2	0
15.00%(w/v) PEG 3350	0	2/85/2	2/75/3	4/10/3
20.00%(w/v) PEG 3350	0	2/100/3	5/60/3	10/20/3
30.00%(w/v) PEG 3350	20/40/2	2/100/3	6/110/3	12/80/3

Results are listed as: the number of crystals per drop / the longest crystal dimension (μm) / the number of drops out of the triplicate that had crystals. † Control experiments use the same solution for drop and reservoir.

Crystals were obtained for each protein with both NaCl and PEG 3350 as SRS precipitant. In some cases, but not all, crystallization success appears to be relatively independent of precipitant concentration in the reservoir solution. There is no clear trend differentiating the traditional and SRS methods in crystal size or number.

3.3.2. Effects of Non-Aqueous Volatile Components in the Crystallization Drop

The failure to obtain crystals when using $(NH_4)_2SO_4$ as SRS precipitant most likely results from the fact that NH_4^+ is in equilibrium with NH_3, and since NH_3 is volatile it can be transferred to the crystallization drop (Mikol *et al.*, 1989). To test if the same is true when $(NH_4)_2SO_4$ is present in the crystallization solution and not in the reservoir solution we attempted to crystallize glucose isomerase using a third known crystallization condition (19% $(NH_4)_2SO_4$ with 0.1 M Hepes at pH 7.2 Rigaku MSC, http://www.rigakumsc.com/protein/crystallization.html). Crystals were obtained using the traditional vapor diffusion method but not when NaCl or PEG 3350 were used as SRS. We also tried using $(NH_4)_2SO_4$ as SRS for this condition and, even though $(NH_4)_2SO_4$ was now present in both the drop and reservoir, no crystals were obtained. This may be due to pH differences, because NH_3 would diffuse towards the solution with the lowest pH where it is captured by protonation to the non-volatile NH_4^+ state. Our experiments suggest that $(NH_4)_2SO_4$ is not suitable for the SRS method. More experiments will be needed to test if this effect is seen for all volatile reagents or only when the reagent also acts as a base or acid.

3.3.3 Crystal Screening

The most interesting application of the SRS method is to use it for crystallization screening where it has the potential to increase the set-up efficiency of large numbers of experiments. To compare the success rate of the SRS method against the traditional vapor diffusion method we screened glucose isomerase, lysozyme, and thaumatin with the PEG ion (Hampton), Cryo I (Emerald BioSystems), and a home-made salt/pH grid screen. Because all PEG-ion screen conditions contain 20% PEG 3350 plus 200 mM of various salts as additive, we used 20% PEG 3350 with 200 mM NaCl as the SRS. This ensures that the dehydrating strength of the SRS is very similar to the individual crystallization

solutions of the screen. The Cryo I screen uses a great variety of precipitant types and concentrations and therefore no single reservoir solution can match their dehydrating strength. In addition, 66% of conditions contain a potentially volatile component that would be absent in the SRS. Nevertheless, we again used 20% PEG 3350 plus 200 mM NaCl (SRS A in Table 3.2), as well as a simpler SRS containing just 20% PEG 3350 (SRS B in Table 3.2). As a final test we screened with a home-made salt grid screen using 1.25 M NaCl as the SRS. 1.25 M NaCl was chosen because results from Table 3.1 suggest it had wide applicability. The results of these experiments are presented in Table 3.2.

Table 3.2 Comparison of the SRS and traditional methods in crystal screens.

	Lysozyme	Glucose isomerase	Thaumatin
PEG/Ion SRS	3/2/20/**25**	4/0/10/**14**	7/6/12/**25**
PEG/Ion traditional	5/2/19/**26**	2/4/12/**18**	4/8/12/**24**
PEG/Ion overlap†	20	11	15
Cryo I SRS A	0/0/9/**9**	0/1/3/**4**	0/0/8/**8**
Cryo I SRS B	0/1/2/**3**	0/0/5/**5**	1/1/22/**24**
Cryo I traditional	0/0/6/**6**	1/0/7/**8**	3/2/7/**12**
Cryo I overlap†‡	4	3	6
Salt grid SRS	1/0/79/**80**	4/0/6/**10**	7/0/20/**27**
Salt grid traditional	1/4/52/**57**	0/1/6/**7**	1/0/19/**20**
Salt grid overlap†	56	1	16

The first three numbers indicate the number of conditions that gave crystals in 1, 2, or 3 drops of the triplicate experiments, respectively. The last number (in bold print) is the sum total. The Cryo I and PEG Ion screens each had 48 conditions and the salt grid screen had 96. † Number of conditions where both methods gave crystals. ‡ The SRS used in this calculation was SRS A.

The traditional and SRS methods have very similar success rates when using the PEG-ion screen. The small differences do not appear to be significant given the observed variation amongst the triplicate experiments. However, despite the similarity in crystallization success, the set of crystallization drop conditions that gave hits in the traditional and SRS

methods are not entirely identical. For lysozyme and glucose isomerase only 80% of the successful conditions gave crystals with both methods. For thaumatin this number was just 60%. Thaumatin also shows the largest variation amongst the triplicate experiments indicating that the lack of correlation between the traditional and SRS method may be due to the random variations inherent in crystallization experiments. However, even robust conditions that gave crystals in each of the three triplicates using one method did not always give crystals with the other method. For the Cryo I screen the success rates of the traditional and SRS methods are more variable as might be expected given the broad range of precipitant strengths and volatile components. For lysozyme and thaumatin there is considerable difference when using 20% PEG 3350 plus 200 mM NaCl (SRS A) or just 20% PEG 3350 (SRS B) as reservoir solution. This result is not surprising given that a 130 mM concentration of NaCl approximates a 20% (w/v) concentration of PEG (Arakali *et al.*, 1995) indicating that the 200 mM NaCl has a significant effect on the dehydrating strength. Interestingly, for both lysozyme and thaumatin one SRS choice outperforms the traditional method. The success rates for glucose isomerase do not differ significantly for the two SRS choices and both are somewhat below the traditional method. We were interested to see if conditions with volatile components in the Cryo I screen have a reduced relative success rate in the SRS mode. We found that for lysozyme, glucose isomerase, and thaumatin the percentage of successful crystallization conditions that contained a volatile component was 66%, 100%, and 63%, respectively. Since 66% of conditions in the Cryo I screen contain a volatile component, there does not appear to be a bias against these conditions when using the SRS method. Finally, for the salt/pH grid screen our choice of 1.25 M NaCl as SRS outperformed the traditional method for all three proteins.

3.3.4 SRS Crystallization Plate

The traditional vapor diffusion method uses identical crystallization and reservoir solutions. This requires that each experiment be set up in individual sealed chambers. The SRS method makes it possible to use a single reservoir solution for many different crystallization experiments. We have exploited this feature in a novel plate design that

makes it possible to incubate many crystallization solutions in a single open chamber. We have named this open chamber plate type, the SRS plate. To test the practicality and crystallization behavior in the SRS plate we have constructed a prototype (Figure 3.1).

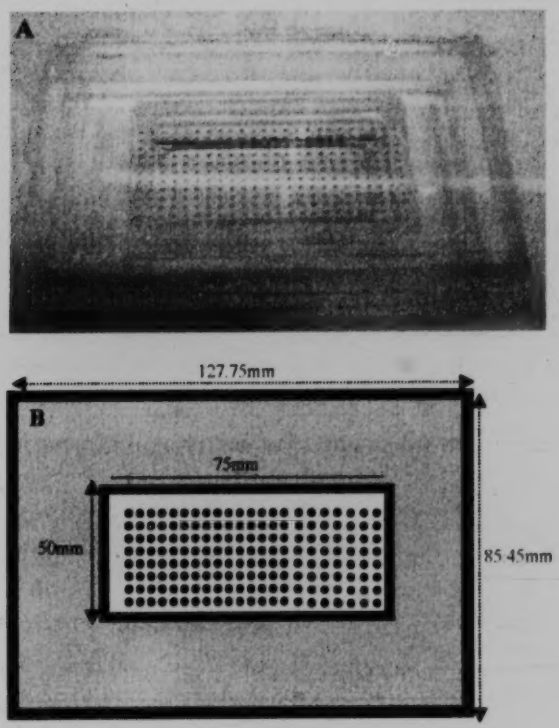

Figure 3.1 SRS Crystallization Plate

A) Prototype SRS crystallization plate. The 288 200 nl crystallization drops are colored with red food dye for visibility. Drops are spaced on a 3 by 3 mm grid. B) Schematic representation of the top view of the prototype plate. Bold lines represent vertical walls of both the plate and central pedestal. Solid grey color indicates plastic bottom and solution in the reservoir. Note that there is no plastic bottom or reservoir solution below the pedestal.

Our prototype open plate has a single chamber in a standard SBS format. The reservoir is filled with a volumetric pipette and the sitting drop platform is made from a single piece of glass with dimensions of 75 x 50 mm^2. This dimension is slightly larger than the 36 x 72 mm^2 taken up by 8 rows and 4 columns of a standard 96-well plate. To close the plate

we use a lid that is also made of glass. The seal between both glass plates and the plastic plate is made with vacuum grease. In our experiments we have dispensed 288 drops of 200 nl each using 3 mm spacing between drops (see Figure 3.1). To compare the success rate of SRS experiments in this plate against a traditional plate, we screened glucose isomerase, lysozyme, and thaumatin with the PEG ion (Hampton Research), and the Cryo I (Emerald Biostructures) screens. The plates compared were the Greiner CrystalQuick conical bottom 96 well plate and prototype SRS plates. We used 20% PEG 3350 with 200 mM NaCl as the SRS in all experiments. Triplicate experiments were carried out in each well of the CrystalQuick plate. Ninety-six triplicate experiments were set up in each SRS plate. The results of these experiments are presented in table 3.3.

Table 3.3 Comparison of the SRS and traditional plates using crystal screens.

	Lysozyme	Glucose isomerase	Thaumatin
Traditional plate	29/8/**37**	19/4/**23**	44/10/**54**
SRS plate	32/7/**39**	23/5/**28**	40/15/**55**

[1] **The first two numbers indicate the number of crystal successes in the PEG ion screen, and the Cryo I screen, respectively, and the last number (in bold print) is the sum total.**

The screening success rates for the SRS and traditional plates are very similar and the differences do not appear to be significant given the observed variation amongst the triplicate experiments.

3.3.5 Optical Quality of SRS Plates

One of our motivations to make the SRS plates is to obtain the best possible imaging of crystallization drops. Design goals were a completely homogeneous background, no crystallization drop well that can distort drop shape and position, and no birefringent

material in the optical path. We have accomplished this by using glass for both the sitting drop surface and the lid. Figure 3.2 shows images of drops using the CrystalQuick and SRS plates, with and without polarization.

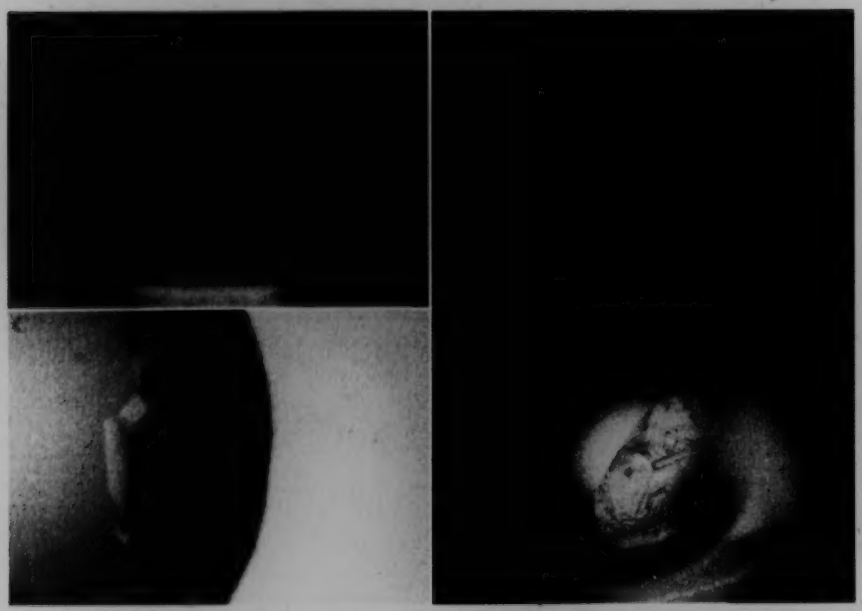

Figure 3.2 Comparison of Crystals in SRS and Commercial Plates.

A) Polarized light image of thaumatin crystals in a flat bottom CrystalQuick plate. B) Polarized light image of thaumatin crystals in a SRS plate at low magnification. C) Polarized light image of glucose isomerase crystals in a SRS plate at high magnification. D) Unpolarized image of a crystallization drop in a conical bottom CrystalQuick plate.

The figure 3.2 images show several common crystallization drop and crystal visualization problems caused by current crystallization plates of various brands. Concave wells have a lens effect and drops often cling to the sides (figure 3.2D) obscuring potential crystals from being imaged. Flat bottom wells give better drop shape (figure 3.2A) but drops with larger volume or lower surface tension often touch and draw up the well edge. Most common plates use plastics that are birefringent and give strong colored background in polarized light (figure 3.2A). In addition, production methods frequently create visible lines in the plastic of the plate (figure 3.2A) that would give a false positive during

automatic classification using edge detection (Spraggon *et al.*, 2002). On SRS plates drops cannot creep up the crystallization drop well edges because the crystallization drops are dispensed on a large flat surface. In addition, both large and small drops can be accommodated by simply selecting appropriate drop spacing, something that is not possible in conventional plates. Drop shape can be further adjusted by using either plain glass or siliconized glass for the drop platform. The use of glass gives optical qualities that are highly favorable. There are no lines or other visual artefacts due to the moulding process of plastics and unlike most plastics the use of glass allows us to use cross polarizers (Figure 3.2B and C). Finally, this type of plate can be recycled by cleaning or replacing the glass slide that forms the sitting drop platform.

3.4 Discussion

3.4.1 Choice of SRS Precipitant

The main requirements of the SRS method reservoir solution is that it has appropriate dehydrating strength and does not contain volatile components. In our experiments, an ammonium sulphate reservoir solution failed to give crystals for all three test proteins. This is most likely due to the ammonium cation, which becomes volatile when it is deprotonated to ammonia. Ammonium sulfate has been used successfully as SRS (Newman, 2005) but the success rate was lower than with other SRS precipitants. In general, we believe that ammonium sulfate is not a good SRS although it may actually be beneficial in isolated cases where, for instance, pH change facilitates crystallization (Mikol *et al.*, 1989; McPherson, 1992). PEG 3350 and NaCl were both successful SRS precipitants in our experiments and other non-volatile precipitants would most likely also be beneficial choices. We prefer the use of NaCl since it is a common, safe, and inexpensive reagent that gives non-viscous solutions.

3.4.2 Effect of Volatile Reagents in the Crystallization Drop

Many crystallization experiments use reagents that are highly or moderately volatile. For instance, 66% of the conditions in the Cryo I screen contain components such as 2-methyl-2,4-pentanediol (MPD), ethyleneglycol, ethanol, or iso-propanol. Because the SRS lacks the volatile reagent, transfer from the drop to the reservoir should occur. Nevertheless, the traditional and SRS methods gave equivalent success rates for crystallization with the Cryo I screen (Table 3.2). In a similar study using lysozyme and Crystal Screen HT (Hampton research), which also contains many conditions with volatile components, the SRS method actually outperformed the traditional method (Newman, 2005). Apparently, the presence of volatile components in crystallization screens in not a problem for the SRS method. Several explanations, or most likely a combination thereof, probably play a role in the outcome: i) the amount of the volatile component that is transferred to the reservoir within the duration of the experiment may be too small to affect the results, ii) the concentration of the volatile component is not always an important parameter, or iii) the transfer of the volatile reagent may have affected some experiments in a positive and others in a negative fashion, without affecting the overall success rate. As discussed above, conditions with ammonium salts may affect crystallization success but it remains to be determined how general this effect is and if it occurs mostly for reagents that can act as an acid or base. Importantly, our results and those reported by Newman (2005) suggest that the presence of volatile components do not appear to reduce the success rate in commonly used crystallization screens, which is the most important criterion for crystallization screening. As we get more experience with the SRS method we may need to replace some of the volatile reagents in our screens or separate out those conditions into a screen to be set up by the traditional method.

3.4.3 Choice of Reservoir Solution Dehydrating Strength

In the traditional method, the mixing of protein sample with reservoir solution assures that the crystallization drop always has a lower dehydrating strength than the reservoir. In

the SRS method this relationship between the dehydrating strength of the drop and reservoir solution has been uncoupled. This raises the issue of how to select a SRS with optimal dehydrating strength. For the PEG-ion screen, the dehydrating strength of all conditions is rather similar since they all contain 20% PEG 3350 plus 200 mM of a salt. Our SRS choice of 20% PEG 3350 plus 200 mM NaCl should therefore, closely match the dehydrating strength of each condition. This is reflected in the similar success rates for the SRS and traditional methods. In most other screens the dehydrating strength varies greatly between the different conditions. We have quantified this by measuring the water activity for many conditions of the Cryo I screen (results shown in chapter 3). This information could be used to group conditions into clusters of comparable water activity that can then be equilibrated together against a SRS with comparable dehydrating strength. However, our results with the three test proteins indicate that it may not be desirable to match the water activity of the SRS to that of the crystallization solution. Indeed, for both lysozyme and thaumatin we found a SRS choice that performed better than the traditional method for the PEG-ion and Cryo I screens. The same was true for our salt/pH grid screen. In a recent study with lysozyme the benefit of the SRS method over the traditional method was even more pronounced, perhaps because a range of reservoir solution concentrations were tested for optimal crystallization success with a specific screen (Newman, 2005).

Our results and the work by Newman, suggest that the water activity of the reservoir solution is an important parameter for crystallization success. In other words, there may be an optimal water activity for protein crystallization and the optimal value likely depends on the target protein and its concentration. This provides a rationale for the observation that the SRS method can outperform the traditional screen because in the latter many conditions will have a water activity far from the optimal value whereas in the SRS method all conditions will reach the same equilibrium water activity that is dictated by the reservoir solution. It has been previously suggested that there is a window of protein-protein interaction energies that is suitable for protein crystallization (McPherson, 1999). Since water from the protein hydration layer is released upon formation of protein contacts, the interaction energy should be a function of water

activity. Although the net interaction energy is the sum total of many different contributions, there may be a window of water activity that is most likely to yield a net interaction energy that is suitable for crystallization. In that case, water activity would be an important predictor of crystallization probability. Unfortunately, the water activity of crystallization solutions has not been considered a relevant parameter and is therefore not normally reported. We have started to measure the water activity for commercial crystallization screen solutions as well as commonly used crystallization precipitants. If we can use this information to calculate an approximate water activity for solutions of given composition then we can use the existing protein crystallization databases to search for correlations between water activity and crystallization success.

3.4.4 Practical Advantages of SRS Plates

The SRS plate described here is a prototype to show that crystallization success rates of such plates are equivalent to our current multi-chambered plates when using the SRS method. The key advantage is that large flat sitting drop platforms can be created on which crystallization drops can be placed at high density and without visualization artefacts often encountered on traditional plates. The flat surface also allows the user to accommodate drops of various sizes by selecting the appropriate drop spacing. Although we believe that a glass sitting drop substrate has superior optical and drop-shape properties, plates made entirely from plastic would still share many of the benefits and may be more practical in some situations. The design of reservoir wells and crystallization platforms is obviously open to many variations. For our Honeybee crystallization robot we are considering a plate with 3 chambers. In this design, standard microscope object slides can be used as the sitting drop platform. Depending on drop radius we expect to be able to place between 96 and 216 drops per chamber (using 4.5 and 3 mm spacing, respectively). However, this design is not compatible with crystallization robots that require a symmetric 96-well plate layout, for example the Hydra96+1 (Robins Scientific) and Hummingbird (Genomic Solutions) robots. For these robots, plates with 8 or 12 chambers can be made by fusing all wells along each row or column, respectively.

Like other plates that are optimised for high-density crystallization screening, our plates are not ideal to harvest crystals, because during harvesting all drops in the chamber are disturbed. We do not consider this a serious disadvantage as finding promising conditions is the principle goal of crystal screening. These lead conditions will be optimised using more traditional techniques that allow convenient harvesting.

3.5 Conclusions

A shared reservoir solution has been used to a limited extent in the past but without serious evaluation of its potential. In particular, the practical consequences of using a shared reservoir solution in our current high-throughput environment had not been explored. Our comparisons of the traditional and shared reservoir solution methods indicate that the latter performs well and is, if the appropriate precipitant strength is used, perhaps even superior to the traditional method. This finding agrees with even more striking results for lysozyme using the Hampton Research Crystal Screen HT (Newman, 2005). An important and unanticipated outcome is thus that the water activity of the reservoir solution may be an important global crystallization parameter that deserves greater attention in future studies.

3.6 References

Abola, E., Kuhn, P., Earnest, T. and Stevens, R. C. (2000). *Nature Struct. Biol.* **7**, 973-977.

Arakali, S. V., Luft, J. R. and DeTitta, G. T. (1995). *Acta Cryst.* **D51**, 772-779.

Berger, I., Kang, C., Sinha, N., Wolters, M. and Rich, A. (1996). *Acta Cryst.* **D52**, 465-468.

Bodenstaff, E. R., Hoedemaeker, F. J., Kuil, M. E., de Vrind, H. P. and Abrahams, J. P. (2002). *Acta Cryst.* **D58**, 1901-1906.

Chu, K., Vojtchovshy, J., McMahon, B. H., Sweet, R. M., Berendzen, J. and Schlichting, I. (2000). *Nature* **403**, 921-923.

Drenth, J. (1994). *Principles of Protein X-ray Crystallography.* New York: Springer-Verlag.

Dunlop, K. V. and Hazes, B. (2003). *Acta Cryst.* **D59**, 1797-1800.

Durbin, S. D. and Feher, G. (1996). *Annu. Rev. Phys. Chem.* **47**, 171-204.

Forsythe, E. L., Maxwell, D. L. and Pusey, M. (2002). *Acta Cryst.* **D58**, 1601-1605.

Fowlis, W. A., Delucas, L. J., Twigg, P. J., Howard, S. B., Meehan, E. J. and Baird, J. K. (1988). *J. Cryst. Growth* **90**, 117-129.

Hui, R. and Edwards, A. (2003). *J. Struct. Biol.* **142**, 154-161.

Krupka, H. I., Rupp, B., Segelke, B. W., Lekin, T. P., Wright, D., Wu, H. C., Todd, P.

and Azarani, A. (2002). *Acta Cryst.* **D58**, 1523-1526.

Mateja, A., Devedjiev, Y., Krowarsch, D., Longenecker, K., Dauter, Z., Otlewski J. and Derewenda, Z. S. (2002). Acta Cryst. **D58**, 1983-1991.

McPherson, A. (1992). *J. Cryst. Growth.* **122**, 161-167.

McPherson, A., Kuznetsov, Y. G., Malkin, A. J. and Plomp, M. (2003). *J. Struct. Biol.* **142**, 32-46.

McPherson, A. (1976). *Methods of Biochemical Analysis.* Vol 23, pp. 249-345, New York: Academic Press.

McPherson, A. (1999). *Crystallization of Biological Macromolecules,* New York: Cold Spring Harbor Laboratory Press.

McPherson, A. (2004). *Methods* **34**, 254œ265.

Mikol, V., Rodeau, J. L. and Giegé, R. (1989). *J. Appl. Cryst.* **22**, 155-161.

Mortuza, G. B., Haire, L. F., Stevens, A., Smerdon, S. J., Stoye, J. P. and Taylor, I. A. (2004). *Nature* **431**, 481-485.

Newman, J. (2005). *Acta Cryst.* **D61**, 490-493.

Spraggon, G., Lesley, S. A., Kreusch, A. and Priestle, J. P. (2002). *Acta Cryst.* **D58**, 1915-1923.

Pros and cons of cryocrystallography: should we also collect a room-temperature data set?

Published in Acta Crystallogr D Biol Crystallogr. 2005 Jan;61(Pt 1):80-7.

Authors: Dunlop, K.V., Irvin, R., and Hazes, B.

Reproduced with kind permission from the International Union of Crystallography.

Chapter 4: Pros and Cons of Cryocrystallography- Should We Also Collect a Room Temperature Data Set?

4.1 Introduction

The goal of macromolecular crystallography is to determine the physiological three dimensional structure of the macromolecule of interest as accurately as possible. This explains the importance of the resolution of the experimental diffraction data, since it determines to a large extent the accuracy and the level of detail that can be obtained for a given structure. Several technical developments in the past decade have led to improvements in resolution, in particular the construction of third generation synchrotrons and the adoption of cryocooling of crystals (Teng *et al.*, 1990). As a result there are now about 150 structures with a resolution of 1.0 Å or better in the protein databank (Berman *et al.*, 2000) and six of these have a resolution better than 0.8 Å. In these structures, hydrogen atoms can be visualized, and nitrogen, oxygen, and carbon atoms can be differentiated on the basis of electron density (Betzel *et al.*, 2001; Kuhn *et al.*, 1998). In addition, atoms with low occupancy, such as weakly bound water molecules and sparsely populated alternate side-chain conformations, can be observed (Betzel *et al.*, 2001). If the resolution of the data set is exceptionally high, for example crambin at 0.54 Å, electron density maps start to shown the features of non-spherical electron orbitals (Jelsch *et al.*, 2000).

Acquisition of the highest possible resolution data requires a high X-ray dose, which in turn necessitates cryocooling of the crystal. However, despite cryogenic cooling the very bright synchrotron beam will still cause radiation damage leading to increased sample mosaicity, increased temperature factor, and decreased diffraction limit (Burmeister, 2000). More importantly, the radiation damage will induce both global and local changes in the protein structure. Global changes include an increase in unit cell volume (Burmeister, 2000; Leiros *et al.*, 2001; Ravelli and McSweeney, 2000) and rotation and translation of the protein molecules (Leiros *et al.*, 2001; Ravelli and McSweeney, 2000).

Local changes include disulphide bond breakage, and loss of carboxyl, hydroxyl, and methylthiol groups of acidic, tyrosine, and methionine residues, respectively (Burmeister, 2000; Leiros *et al.*, 2001; Ravelli and McSweeney, 2000; Weik *et al.*, 2000). Although this does not normally prevent determination of a high resolution protein structure it must be kept in mind that radiation damage may have caused structural artifacts.

Cryogenic techniques greatly reduce radiation damage but, unfortunately, they can also introduce new artifacts resulting directly from temperature effects or, indirectly, from the addition of cryoprotectants and temperature-induced pH changes. Indeed, decreases in unit cell volume, changes in molecular packing, and perturbation of local structure have been reported (Juers and Matthews, 2001). In a recent theoretical study, Halle (Halle, 2004) proposes that cryogenic crystal structures are not kinetically trapped in a room-temperature equilibrium state. Because crystal cooling is slow, ≥ 0.1 second (Kriminski *et al.*, 2003), proteins adapt to the changing free energy landscape during cooling until their conformation is immobilized when the glass transition is reached at around 200 K (Miyazaki *et al.*, 2000). Halle therefore predicts "substantial cryoinduced shifts of substate populations" where the structure largely reflects the free energy state at the glass transition temperature of the solvent. The effect is expected to be most pronounced at the protein surface, a region of critical functional importance. It is unlikely that cryoinduced artifacts can be prevented because that would require crystal cooling rates near the picosecond timescale of molecular motions. It is therefore important to determine the nature and extent of temperature-induced changes.

We have previously solved a room temperature crystal structure of truncated PAK pilin at 1.63 Å resolution (Hazes *et al.*, 2000). A 0.78 Å data set was subsequently collected at 100 K on beamline SBC-19ID at the Advanced Photon Source (Chicago, U.S.A). Distinct differences in protein and hydration structure were noted and only part of these could be ascribed to the increased resolution. We therefore collected additional data sets to determine the role of resolution, radiation damage and crystal temperature on these differences. In total three cryogenic and three room temperature PAK pilin data sets were collected using 3rd and 2nd generation synchrotrons and rotating anode X-ray sources.

Our results indicate that cryogenic cooling of crystals introduces systematic changes that are large relative to the expected coordinate error. Moreover, although the changes are most pronounced at the protein surface, even the main-chain atoms of residues forming the main secondary structural elements are affected.

4.2 Materials and Methods

4.2.1 Crystallization, Data Collection, and Data Processing

Expression, purification, and crystallization of truncated PAK pilin (residues 29-144) has been previously described (Hazes *et al.*, 2000). Briefly, crystals were grown using the hanging drop vapor diffusion method by equilibrating a 6 µL drop with equal volumes of protein solution (10 mg/mL in doubly distilled water) and precipitant solution (60 % $(NH_4)_2SO_4$, 0.1 M Hepes pH 8.2) against 1 mL of the precipitant solution. Seeding was used to obtain crystals typically sized 0.3 x 0.3 x 0.2 mm^3. All crystals belong to space group $P4_12_12$.

For the cryogenic data sets, fresh crystals were transferred to a solution consisting of 20 % glycerol and 80 % precipitant solution and flash cooled in a nitrogen gas stream at 100 K. Room temperature crystals were mounted in quartz capillaries. For data set 5, the crystal used previously to collect cryogenic data set 3 was returned to standard mother liquor and remounted in a quartz capillary. Data set 1 was processed with D*TREK (Pflugrath, 1999) whereas all others were integrated with MOSFLM (Leslie, 1992) and scaled with SCALA (Collaborative Computational Project, 1994).

Data set 1 was collected as four consecutive batches (Table 4.1). During processing it became clear that significant radiation damage had occurred. The first batch had sustained the least amount of damage and was therefore the preferred source of experimental data. However, due to saturation of the detector no data below 2.5 Å was included in the first batch. To obtain the final data set, scaling and merging was carried out in a nonstandard manner. First, batch 1 was scaled and merged separately. Next, all

four batches were processed together and scaled to batch 1. Finally, the program SFTOOLS (B. Hazes, unpublished results) was used to create data set 1 by taking all reflections from batch 1 and just adding the missing reflections from the scaled set of all data. This minimized the effect of radiation damage without compromising data completeness.

Table 4.1 Batch statistics for the ultra-high 0.78 Å resolution data set (data set 1).

Batch	Max. resolution (Å)	No. images	Exposure (s)	Attenuator
1	0.76	500	5	None
2	0.81	530	10	None
3	1.12	240	5	0.75 mm Al
4	1.64	120	5	0.13 mm Ag

4.2.2 Refinement

The previously published 1.63 Å PAK pilin structure (PDB entry 1DZO; Hazes *et al.*, 2000) was used as the starting model for all refinements. To calculate Rfree statistics, each data set inherited the test set from the 1.63 Å structure. For data exceeding 1.63 Å resolution a random set of 5 % of the new reflections were added to the test set. For all data sets, an initial round of rigid-body refinement was performed with REFMAC (Murshudov *et al.*, 1997) to position the molecule more accurately in the cell. Subsequently, cycles of REFMAC maximum-likelihood positional and B-factor refinement were alternated with manual remodeling using Xfit (McRee, 1999) until R and Rfree-factors converged. For the models based on data sets 1 and 2, riding hydrogens were introduced followed by further positional and anisotropic temperature factor refinement in REFMAC. Finally, SHELXL (Sheldrick and Schneider, 1997) was used to assign occupancies for side chains with multiple conformations. Water molecules were added using 20 cycles of ARP/wARP (Perrakis *et al.*, 1999) after which waters without clear density in a 2Fo-Fc density map and waters not within 2.5 - 3.3 Å of a nitrogen or oxygen atom were manually discarded. This semi-automatic water building method was

chosen to minimize bias so the water molecules in all six structures could be objectively compared. To obtain the final hydration structures, more water molecules were subsequently added based on visual inspection of electron density maps.

4.2.3 Hydrogen Atom Density

Hydrogen atoms were deleted from the final models of high resolution structures 1 and 2. The program PDBSET (Collaborative Computational Project, 1994) was used to apply a random Gaussian shift (r.m.s. = 0.1156 Å; maximum shift = 0.2 Å) to the coordinates to remove model bias. After refinement of the resulting model without hydrogen atoms, Sigmaa weighted difference map Fourier coefficients from REFMAC were read by the program HYDENS (B. Hazes, unpublished results) to quantify the electron density signal at riding hydrogen positions using a classical Fourier summation.

4.2.4 Hydration Shells

The program MCFMAN (B. Hazes, unpublished results) was used to assign hydration shells. The program LSQKAB (Collaborative Computational Project, 1994) was used to superimpose pairs of structures, after which the programs GROUPWATER and H2ODIST (B. Hazes, unpublished results) were used to assign and tabulate groups of conserved water positions. Individual water oxygen molecules were considered to correspond to equivalent positions if they lay within a distance of 1 Å of each other. The same procedure was used to analyse the lysozyme structures (PDB ID, 3LZT and 4LZT; Walsh *et al.*, 1998).

4.2.5 Program availability

SFTOOLS is distributed as part of the CCP4 package (Collaborative Computational Project, 1994). All other unpublished programs are available at (http://eagle.mmid.med.ualberta.ca/highlights.html).

4.3 Results and Discussion

4.3.1 Ultra-High Resolution Structure and Radiation Damage

The crystal structure of PAK pilin has been determined at 1.63 Å resolution using diffraction data collected at room temperature on a rotating anode X-ray source (Hazes *et al.*, 2000). Subsequently, a 0.78 Å resolution data set was obtained at cryogenic temperature on beamline SBC-19ID of the Advanced Photon Source (Chicago, U.S.A). A data set based on all diffraction images gave significantly improved electron density maps. However, the R and Rfree refinement statistics converged at 16.4 % and 17.9 %, respectively, which is considerably higher than comparable ultra-high resolution structures. We anticipated that this was a result of radiation damage and re-refinement against the "batch 1" data set, a 0.78-2.5 Å data set based on the least damaged first 500 images (see methods), reduced R and R-free to 12.7 % and 14.1 %, respectively. To obtain a complete data set with reduced radiation damage we took every reflection present in the 0.78-2.5 Å data set and added only the missing reflections from the more damaged data sets (see methods). The final model was refined against these data yielding R and Rfree factors of 14.2 % and 15.4 %, respectively.

4.3.2 Radiation Damage to the Disulphide Bond

In the 1.63 Å PAK pilin structure a well-defined disulphide bond is observed between Cys129 and Cys142 (Figure 4.1A). In the 0.78 Å structure this disulphide bond is still present but a second conformation of Cys142 Sγis clearly indicated by the electron density (Figure 4.1B). Both cysteines were therefore modeled with a second "B" conformation. The distance between the sulfurs in the "B" conformation is 2.78 Å, which is too long for a disulphide bond and too short for a normal van der Waals interaction. This distance is indicative of a disulphide radical (Carugo and Bordo, 1999). Occupancy refinement in SHELXL (Sheldrick and Schneider, 1997) resulted in occupancies of 0.54 for both sulfurs forming the disulphide bond, and 0.38 and 0.29 for the alternate conformations of Cys129 and Cys142, respectively. The reduction in the total occupancy

for the Sγatoms of Cys129 and Cys142 to 0.83 and 0.92, respectively, suggests that sulfur atoms have been lost due to radiation damage, a phenomenon that has been previously reported (Burmeister, 2000; Leiros *et al.*, 2001). No damage to the disulphide bond has been observed with any of the rotating anode data sets, whilst less damage was seen with data set 2, collected at a 2nd generation synchrotron, or when only the first 500 images of data set 1 were used (results shown in appendix).

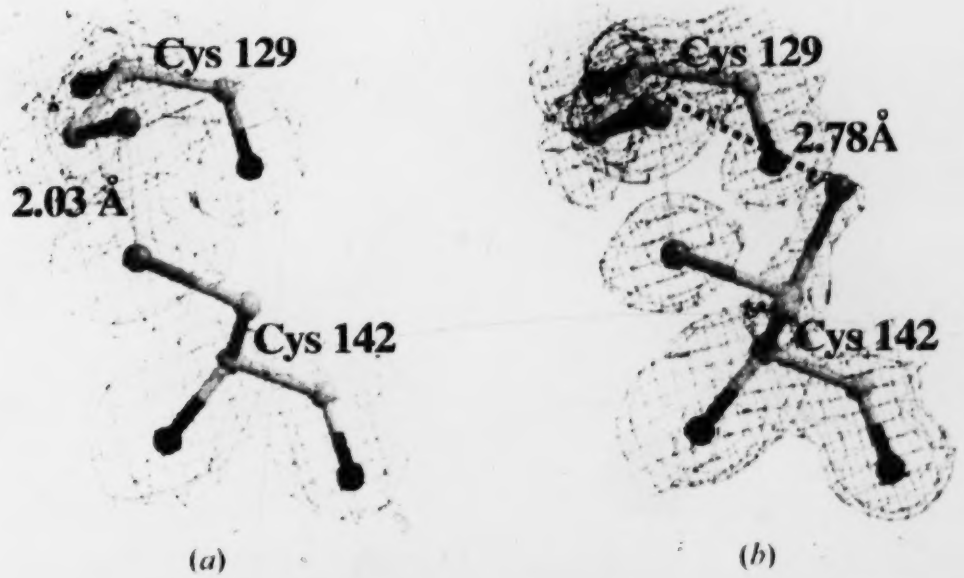

Figure 4. 1 2Fo –Fc electron density for cysteine residues 129 and 142 contoured at 1.5 σ.

(A) Density for structure 4, based on rotating anode diffraction data to 1.63 Å resolution, shows a single conformation of the disulphide bond. (B) Density for structure 1, based on 0.78 Å diffraction data from a third-generation synchrotron source, shows two conformations for the sulfur atom of cysteine 142. Cysteines 129 and 142 have both been given dual side-chain conformations to model the intact and broken disulphide, respectively. This yields a disulphide bond with a typical length of 2.03 Å and two sulfur atoms at a distance of 2.78 Å, representative of a disulphide radical (Weik *et al.*, 2002). The figure was made using Bobscript (Esnouf, 1997) and Raster 3D (Merritt & Murphy, 1994). Carbon, nitrogen, oxygen, and sulfur atoms are displayed in yellow, blue, red, and green, respectively.

4.3.3 Hydrogen Atom Density

In addition to the very local effect of radiation damage on the disulphide bond we also observed a global effect on electron density quality. This was most obvious when looking for weak signals such as experimental density for hydrogen atoms. As hydrogen atoms have only one electron, their detection is highly susceptible to radiation damage-induced noise in the electron density map. When we evaluated the experimental electron density at riding hydrogen positions in a Sigmaa weighted Fo-Fc difference map (see methods), we could only detect 0.76 % of the hydrogens above 2σ when using all data (Figure 4.2A). However, if the difference map was based on the 0.78-2.5 Å data set with less radiation damage then 27.4 % of the riding hydrogens had experimental density above 2σ and 5.7 % had density stronger than 3σ. The latter difference map also showed density for one polar hydrogen, the hydroxyl hydrogen of threonine 95 (Figure 4.2B). In the 0.78 Å crystal structure of *Bacillus lentus* subtilisin 14 % of all riding hydrogens had a signal above 3σ (Kuhn *et al.*, 1998). In hindsight, our attempt to push resolution by using a high X-ray dose probably decreased our ability to detect hydrogen atoms due to increased radiation damage.

(*a*) (*b*)

Figure 4.2 Hydrogen atom density

Fo-Fc electron density maps for threonine 95 in structure 1 contoured at 1.5 σ. To remove model bias, the final structure 1 model was re-refined after removal of the hydrogens and applying a small random perturbation of the coordinates (see methods). (A) Map based on all data shows density for only two hydrogen atoms. (B) Map based on just the data in batch 1, which has less radiation damage, shows six hydrogen atoms. The figure was made using Bobscript (Esnouf, 1997) and Raster 3D (Merritt & Murphy, 1994). Hydrogen, carbon, nitrogen, and oxygen atoms are displayed in white, yellow, blue, and red, respectively.

4.3.4 Temperature-Induced Structural Changes

A comparison of the 1.63 Å and 0.78 Å structures showed differences in multiple side-chain conformations, small differences in general atomic coordinates, and a much larger number of bound water molecules in the high resolution structure. To determine the contributions of resolution, radiation damage and temperature to these differences we collected four more data sets. All data sets are based on fresh crystals except data set five. To determine if cryoinduced changes are reversible, we chose to re-use the crystal that gave cryogenic data set 3 for the room temperature data set 5. Details of the data collection and refinement statistics for all six data sets are given in Table 4.2.

Table 4.2 Diffraction data and refinement statistics.

Data set	1†	2	3	4	5	6
Light source	APS	SRS	Rotating anode	Rotating anode	Rotating anode	Rotating anode
Beamline	19ID	9.6	–	–	–	–
Detector	SBC-2 CCD	ADSC Q210	R-AXIS IV	MAR 300	R-AXIS IV	R-AXIS IV
Temperature (K)	100	100	100	293	293	293
Wavelength (Å)	0.65	0.97	1.54	1.54	1.54	1.54
Maximum resolution (Å)	0.78	0.95	1.51	1.63	1.80	1.80
Unit-cell parameters (Å)						
$a = b$	37.24	37.58	37.48	38.11	38.04	37.81
c	147.86	148.56	147.89	149.78	149.58	149.44
Total observations	1274912 (502237)	1198789	165793	110628	167068	312705
Unique observations	107451 (100725)	67341	16080	14500	10953	10797
R_{merge}‡ (%)	8.2 (80.1/35.3 (33.7)	7.4/12.3	5.0/12.2	4.9/19.0	7.9/49.8	10.9/25.1
$I/\sigma(I)$‡	5.47 (3.38)/3.0.70 (1.07)§	21.41/2.46	23.50/4.09	26.70/10.10	11.59/2.50	14.62/7.22
Completeness‡ (%)	89.8 (84.3)/42.1 (42.4)	75.0/4.5††	91.7/16.3††	99.5/89.6	98.6/76.0	98.6/62.1
Wilson B factor (Å^2)	6.8 (6.1)	7.4	12.2	16.8	19.0	15.0
R factor (%)	14.2	11.3	13.8	14	14.8	14.5
R_{free} (%)	15.4	13.1	16.1	15.8	18.5	17.2
Bond length (Å)	0.019	0.007	0.011	0.009	0.020	0.014
Bond angle (°)	1.99	1.34	1.42	1.34	2.43	1.54
Estimated coordinate error	0.01	0.02	0.07	0.07	0.10	0.07

† Numbers in parentheses are statistics for the first 500 images (batch 1, 0.76-2.5 Å). ‡ Overall and "highest resolution shell" statistics are listed, separated by a forward slash. The highest resolution shells for data sets 1-6 are, 0.78 -0.79, 0.95-1.01, 1.51-1.55, 1.63 -1.72, 1.8-1.85, and 1.8-1.85 Å, respectively. § Average I/σI is around 60 prior to merging, as expected for a strongly diffracting crystal. During merging, σI is scaled up to account for larger than expected variations between symmetry related reflections; most likely at least in part due to radiation damage. This reduces the average I/σI to the low values listed in the table. ¶ Inclusion of data up to 0.78 Å was not based on I/σI, which is unusually low for reasons explained above. R-merge, although typically a less reliable statistic, did indicate useful signal to this resolution. This was confirmed by the fact that calculated amplitudes from a model refined with data to 0.85 Å resolution showed a positive correlation with the observed amplitudes to 0.78 Å resolution. †† Low completeness at high resolution is due to the square shape of the detector. At the shortest allowed distance, only the detector corners contributed to the highest resolution reflections.

4.3.5 Structural Variation in the Main Chain

The truncated PAK pilin contains residues 29 to 144 of the native protein preceded by four residues, 25-28, derived from the expression vector. The main-chain structure of all models is basically identical but superimposing pairs of structures using the program LSQKAB (Collaborative Computational Project, 1994) shows interesting deviations

(Table 4.3). For the room temperature structures, the r.m.s. deviations between residues that form the main secondary structural elements (core residues) are a little less than 0.05 Å and this increases to between 0.07 and 0.10 Å when all residues are used for the analysis. For the cryogenic structures these numbers range from 0.05 to 0.08 Å for the core residues and 0.09 to 0.13 Å when considering all residues. Except for the comparison of structures 1 and 2, these values are not significantly different from the expected coordinate error (Table 4.2) and the higher values for the cryogenic structures may reflect lower reproducibility of the cryogenic cooling step. Significantly higher values are obtained when comparing the room temperature structures with the cryogenic structures, 0.18 to 0.23 Å for the core residues and 0.20 to 0.27 Å for all residues (boxed regions in Table 4.3). The increase is not simply due to the typical cryoinduced 1-2 % contraction of proteins (Juers and Matthews, 2001), because computational contraction of the room temperature structures improved the r.m.s. values by only ~0.04 Å (with an optimal contraction factor of 1 %). Clearly, cryoinduced shrinking of the protein does not explain the large r.m.s. values between room temperature and cryocrystal structures. The high r.m.s. values also do not appear to be due to radiation damage. Although structure 1 (highest radiation dose) has the highest r.m.s. values, even structure 3 (with no observable radiation damage) shows high r.m.s. values when compared to the room temperature structures. Therefore any radiation damage effect must be small. Instead, we conclude that cryogenic cooling causes small but significant systematic changes that are not just rigid body rotation, translation, or scaling effects. The effects are larger in loop regions but can be readily detected in even the core residues of the protein. These cryoinduced changes are reversible, as structure 5 behaves like a typical room temperature structure even though a cryocooled crystal was thawed to collect this data set.

Table 4.3 Structural deviations between pairs of PAK structures[i].

Temperature (K)		100	100	100	293	293	293
	Structure	1	2	3	4	5	6
100	1	—	0.079	0.069	**0.232**	**0.222**	**0.214**
100	2	0.130	—	0.047	**0.192**	**0.188**	**0.178**
100	3	0.088	0.107	—	**0.204**	**0.195**	**0.185**
293	4	**0.267**	**0.238**	**0.229**	—	0.041	0.049
293	5	**0.257**	**0.232**	**0.219**	0.092	—	0.047
293	6	**0.239**	**0.221**	**0.204**	0.069	0.103	—

[i] Pairs of PAK structures were superimposed using LSQKAB (Collaborative Computational Project, 1994) and r.m.s. deviations for all superimposed main-chain atoms, including Cβ, are listed. Results for superimposing all residues are given below the diagonal. Results for superimposing only the core residues, those that form the α-helix and the main four β-strands, are given above the diagonal. Bold numerals highlight comparisons between structures determined at different temperatures.

4.3.6 Structural Variation in the Side Chains

In our comparison of the 1.63 Å and 0.78 Å structures we noticed changes in side-chain conformations that were not readily explainable by increased resolution or the presence of radiation damage. To determine whether this reflects crystal to crystal variation or a more systematic cryoinduced change, we compared the six PAK structures. As Cys129 and Cys142 have a conformational change that is clearly induced by radiation damage these residues were not considered in this analysis. We analyzed 13 residues with dual side-chain conformations in at least one of the six structures and one residue, Gln136, with a single side-chain conformation that was not identical in all structures. The results are summarized in Table 4.4.

Table 4.4 Multiple side-chain conformations of the six PAK structures.

	Glu27	Arg30	Ser31	Glu49	Ser52	Arg53	Val57	Thr84	Lys88	Thr99	Met104	Lys110	Lys112	Gln136
1	*AB*	*A*	*A*	*AB*	*AB*	*AB*	*AB*	*AB*	*AB*	*A*	*AB*	*AB*	*AB*	*A*
2	*AB*	*A*	*A*	*A*	*AB*	*AB*	*AB*	*AB*	*AB*	*A*	*AB*	*A*	*AB*	*A*
3	*AB*	*A*	*A*	*AB*	*AB*	*AB*	*AB*	*AB*	*AB*	*A*	*AB*	*A*	*AB*	*A*
4	*A*	*AB*	*AB*	*B*	*AB*	*A*	*BC*	*AB*	*C*	*AB*	*AB*	*A*	*AB*	*B*
5	*A*	*A*	*AB*	*B*	*AB*	*A*	*BC*	*AB*	*C*	*AB*	*AB*	*A*	*AB*	*B*
6	*A*	*A*	*AB*	*B*	*AB*	*A*	*BC*	*AB*	*C*	*AB*	*AB*	*A*	*AB*	*B*

The letters A, B, and C indicate distinct side-chain rotamers.

Four out of 13 residues with multiple conformations (Ser52, Thr84, Met104, and Lys112) display the same conformations in all six structures. Two residues (Arg30 and Lys110) have dual side-chain conformations in only one of the 6 structures. This may reflect individual differences between crystals. In all other cases side-chain conformations appear to be strictly correlated with temperature (Figure 4.3). Clearly, radiation damage does not play a role in side-chain conformation as structure 3 behaves the same as structures 1 and 2. It is also unlikely that resolution explains the difference as structure 3 is based on data with a resolution that is only slightly higher than structure 4. More importantly, residues Ser31 and Thr99 have dual conformations at room temperature but not in the cryogenic structures and residues Val75, Lys88, and Gln136 have different conformations in the room temperature and cryogenic structures. These observations indicate actual systematic differences as a function of temperature. An example of different side-chain conformations is given for Glu27 in Figure 4.3.

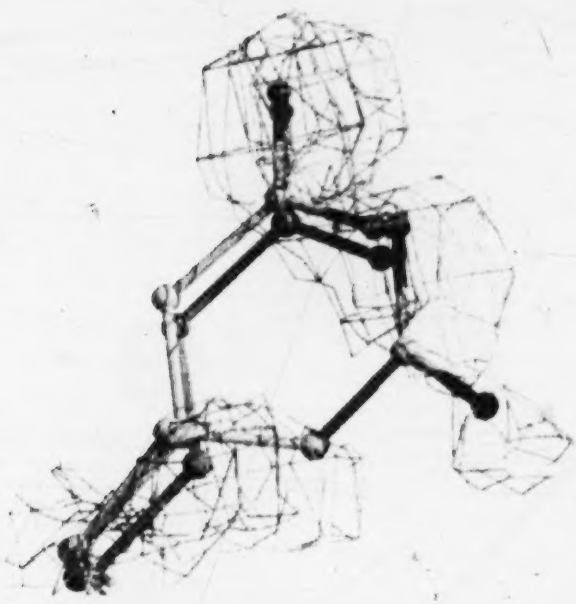

Figure 4.3 Temperature related alternate side chain conformations.

Fo-Fc density for Glu27 contoured at 1σ. Density for structure 3 (yellow) shows two Glu27 side chain conformations. Density for structure 4 (green) shows only one side-chain conformation. Structure 3 and 4 are representative of all cryocooled and all room temperature structures, respectively. The figure was made using Xfit (McRee, 1999) and Raster 3D (Merritt & Murphy, 1994).

Six residues were modeled with incomplete side chains in at least one of the six structures due to a lack of electron density. These are all long hydrophilic surface residues, including four of the eleven lysine residues. The data presented in Table 4.5 show that the room temperature crystal structures have a significantly larger number of incomplete side chains than the cryostructures. The lower resolution of the room temperature structures may contribute to the inability to observe weakly occupied side-chain conformations. However, the fact that the number of incomplete side chains in structure 3, which has only slightly higher resolution than the room temperature structures, is the same as the ultra-high resolution structure 1, suggests that temperature again plays a dominant role. Differential occupation of conformational states depends on both the free energy difference of the states and the temperature, as described by the

Boltzmann distribution law. As a consequence, for a given free energy difference, higher temperature will increase structural disorder. This can lead to a situation where atoms can no longer be detected in the electron density map. Tables 4.4 and 4.5 both show that the side-chain conformations of structure 5 match those of the other room temperature structures, indicating that cryoinduced effects on side-chain conformations are also reversible.

Table 4.5 Incomplete side chains for the six PAK structures[1].

	Arg30	Lys68	Lys88	Lys110	Lys128	Glu135
1	√	√	√	C^{δ}	√	√
2	√	C^{ε}	√	C^{δ}	√	√
3	√	√	√	C^{δ}	√	√
4	C^{δ}	√	C^{γ}	√	C^{γ}	C^{γ}
5	C^{β}	C^{δ}	C^{γ}	C^{γ}	C^{γ}	C^{β}
6	C^{β}	C^{δ}	C^{γ}	C^{γ}	C^{γ}	C^{β}

[1] **Atom names indicate the last visible atom in the side chain. √ indicates that the side chain is complete.**

4.3.7 Differences in Hydration Structure

Water molecules are important for many properties of proteins, including molecular recognition, enzyme activity, stability, and folding. Unfortunately protein hydration is often the most difficult aspect of protein structure to determine by X-ray diffraction due to several technical problems. First, protein-protein interactions in crystal packing interfaces affect the hydration structure of the protein surfaces near those interfaces. Secondly, water molecules tend to be more difficult to model than protein atoms as their electron density signal is normally weaker due to a broader positional distribution (Carugo, 1999). Third, their refinement is more difficult since, without covalent bonding, there are no strong bond length and bond angle restraints to help define the atomic position. The severity of these last two issues is resolution dependent and the

completeness of the hydration model in crystal structures therefore decreases at lower resolution. Indeed, Carugo and Bordo (1999) who studied 873 room temperature and 33 low temperature protein structures at a wide range of resolutions concluded that the number of modeled water molecules depended primarily on the resolution of the structure, while the temperature of the crystal during data collection had only a minor effect. In contrast, in a study of the hydration of several 1.8 Å bovine β-trypsin structures, Nakasako (1999) reported that cryogenic structures contained 1.5 to 2.1 times more visible bound water molecules than room temperature structures. In the analysis of our six structures we observed that hydration levels of cryogenic structures were 1.9 to 2.7 times that of the room temperature structures (Table 4.6). This is even notable when comparing structures 3 and 4 which are of comparible resolution. Resolution may have a small effect as hydration level correlates with resolution for the cryogenic structures. However, our results, combined with the observations for bovine β-trypsin (Nakasako, 1999) and lysozyme (see below), indicate that cryocooling of crystals leads to a large increase in the number of detectable water molecules with a smaller effect for resolution. Resolution may be more important for lower resolution structures, such as those included in the study by Carugo and Bordo (1999).

Table 4.6 Number of water molecules and water shell distribution for the six PAK structures.

Structure	1	2	3	4	5	6
Shell 1	154 (69%)	151 (71%)	139 (71%)	90 (88%)	72 (88%)	84 (86%)
Shell 2	65 (29%)	56 (26%)	50 (25%)	12 (12%)	9 (11%)	14 (14%)
Shell 3	5 (2%)	5 (2%)	8 (4%)	0	1 (1%)	0
Total	224	213	197	102	82	98

The percentage of water molecules in each shell is given in parentheses.

To extend our analysis, we grouped the water molecules in shells. First shell waters have at least one direct hydrogen bond to a protein atom. Second shell waters have at least one direct hydrogen bond to a first shell water and third shell waters only make hydrogen

bonds to second shell waters. Table 4.6 shows that for all structures the majority of waters (69-88 %) are located in the first shell, a lesser number in the second shell (11–29 %), and waters are rarely located in the third shell (0-4 %). Given that cryocooling had a large effect on the number of detectable waters, the effect on first and second shell waters was analyzed. This comparison shows that cryocooling causes a much greater increase in second shell water numbers (7.1 fold) as compared to first shell water numbers (1.8 fold). To determine whether these hydration effects occur for other proteins we compared two lysozyme structures (PDB ID 3LZT and 4LZT; 22). 3LZT was refined against cryogenic data to 0.92 Å resolution and 4LZT was based on room temperature data to 0.95 Å resolution. Both data sets were derived from crystals with the same unit cell and space group. The number of water molecules observed in the cryogenic structure was 1.84 times that of the room temperature structure and the percentage of waters in shells 1, 2, and 3 was 88 %, 11 %, <1% for the room temperature structure and 77 %, 21 %, 2 % for the cryogenic structure. These numbers are very similar to our observations for PAK pilin and once more demonstrate that cryocooling has a large impact on the extent of hydration in crystal structures.

After showing that cryocooling significantly increases the number of modeled water molecules, we wished to determine whether the room temperature water positions were conserved at cryogenic temperature. The results are listed in Table 4.7.

Table 4.7 Structurally equivalent water molecules among pairs of PAK structures.

	1	2	3	4	5	6
1	–	138/46/3/187	130/37/2/169	80/10/0/90	63/9/0/72	79/11/0/90
2	91/82/60/88	–	138/39/2/179	84/9/0/93	67/8/0/75	81/11/0/92
3	94/74/25/86	99/78/25/91	–	77/7/0/84	60/6/0/66	74/10/0/84
4	89/83/0/88	93/75/0/91	86/58/0/82	–	69/7/0/76	81/11/0/92
5	88/100/0/88	93/89/0/91	83/67/0/80	96/78/0/93	–	68/7/0/75
6	94/79/0/94	96/79/0/94	88/71/0/86	96/92/0/94	94/78/0/91	–

The number of equivalent waters is listed above the diagonal. Percentages are listed below the diagonal. The four values represent the results for shell1/shell2/shell3/total.

A very high percentage of water positions are conserved between all six structures (80-94 %) suggesting that there is a well defined subset of bound waters shared by all structures. This subset largely corresponds to the waters observed in the room temperature structures. Within this subset, the extent of conservation between all room temperature structures depends on the water shell. Conservation is higher for first shell waters, averaging 92 %, than for second shell waters, averaging 79 %. Waters that are observed in all room temperature structures are to a very large extent also present in the cryogenic structures; only 3 water molecules shared by all room temperature structures are missing from the cryogenic structures. In contrast, 73 waters that are found in all cryogenic structures are not present in any of the room temperature structures (Figure 4.4).

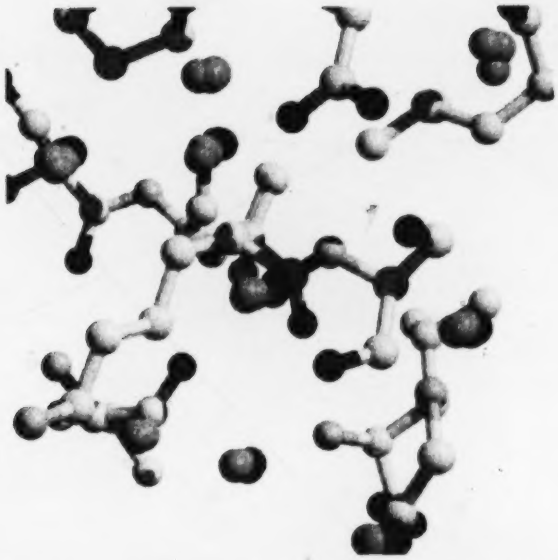

Figure 4.4 Superimposed Cryocooled and Room-Temperature Water Molecules

Ball and stick model showing superimposed cryocooled (green) and room-temperature (dark blue) water molecules for all six PAK structures within a small section of the PAK protein. The structure shows that waters observed in room temperature structures are also present in cryocooled structures but not *visa versa.* The figure was made using Xfit (McRee, 1999) and Raster 3D (Merritt & Murphy, 1994).

Three distinct effects may contribute to cryoinduced changes in hydration. 1) The higher resolution of cryogenic data allows the detection of more poorly defined waters. However, the equal resolution of the lysozyme structures and the fact that we see most cryoinduced changes already in structure 3 at 1.51 Å resolution, suggests that this effect is small. 2) Reduced thermal motion and/or a smaller entropic penalty for ordering water at lower temperature may reduce positional disorder and raise the electron density to detectable levels. Interestingly, the average B-factor for the cryogenic waters that are not observed in room temperature structures is 35.72 $Å^2$. This is higher than for the cryogenic waters that are also present at room temperature (28.47 $Å^2$), indicating that the latter are more strongly localized hydration sites. 3) Changes in the free energy landscape due to structural rearrangements or direct temperature effects on the force field may change the location or existence of water binding sites. This does not explain the systematic cryoinduced increase in hydration level but could account, for instance, for the observation of the three water molecules that are only found in the room temperature structures.

4.4 Conclusions

Brilliant synchrotron X-ray sources combined with cryocooling have made it possible to solve protein structures from crystals that were previously considered either too small or diffracting too weakly. Even for crystal structures that can be solved with conventional X-ray sources the use of cryogenic data collection at a high brilliance synchrotron source significantly improves resolution and data quality. It is however, important to remain aware of potential drawbacks. The occurrence of radiation damage to specific chemical groups, in particular disulphide bonds, is now well-documented. In contrast, the effect of cryocooling has received much less attention, perhaps because it is less conspicuous and only becomes clear in direct comparisons between equivalent structures solved at ambient and cryogenic temperatures. Our work using three data sets at each temperature demonstrates that there are indeed consistent cryoinduced structural effects. There are small coordinate shifts throughout the protein but these are not likely to impact biological interpretations. More importantly, discreet changes are observed in hydration structure and the conformation of side chains, especially surface exposed side chains that already display some disorder. In a few cases, bound waters and side-chain conformations seen at room temperature no longer exist in the cryogenic structures. However, the predominant effect is a reduced disorder at cryogenic temperature which results in fewer disordered side chains and more extensive hydration. Since most biological processes occur at the protein surface it is important to be aware that the physiological side-chain conformation and hydration structure may differ from the cryogenic crystal structure. Basically, the interpretation of protein surface features in a cryogenic structure needs the same caution as has traditionally been applied to the interpretation of surface features in crystal packing interfaces. It is our opinion that in the great majority of cases the benefits of cryogenic synchrotron diffraction data outweigh the drawbacks, and often there may be no choice. However, our results indicate that for crystals with sufficient diffraction power it is advisable to also collect a room temperature data set.

4.5 References

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000) *Nucleic Acids Res.* **28,** 235-242.

Betzel, C., Gourinath, S., Kumar, P., Kaur, P., Perbandt, M., Eschenburg, S. & Singh, T. P. (2001) *Biochemistry* **40**, 3080-3088.

Burmeister, W. P. (2000) *Acta Cryst.* **D56**, 328-341.

Carugo, O. (1999) *Protein. Eng.* **12,** 1021-1024.

Carugo, O. & Bordo, D. (1999) *Acta Cryst.* **D55,** 479-483.

Collaborative Computational Project, Number 4. (1994) *Acta Cryst.* **D50**, 760.

Esnouf, R. (1997) *J. Mol. Graph.* **15**, 133-138.

Halle, B. (2004) *Proc. Natl. Acad. Sci.* USA **101,** 4793-4798.

Hazes, B., Sastry, P. A., Hayakawa, K., Read, R. J. & Irvin, R. T. (2000) *J. Mol. Biol.* **299,** 1005-1017.

Jelsch, C., Teeter, M. M., Lamzin, V., Pichon-Pesme, V., Blessing, R. H. & Lecomte, C. (2000) *Proc. Natl. Acad. Sci.* USA **97,** 3171-3076.

Juers, D. H. & Matthews, B. W. (2001) *J. Mol. Biol.* **311**, 851-862.

Kriminski, S., Kazmierczak, M. & Thorne, R. E. (2003) *Acta Cryst.* **D59**, 697-708.

Kuhn, P., Knapp, M., Soltis, S. M., Ganshaw, G., Thoene, M. & Bott, R. (1998) *Biochemistry* **37**, 13446-13452.

Leiros, H. K., McSweeney, S. M. & Smalas, A. O. (2001) *Acta Cryst* **D57**, 488-497.

Leslie, A. G. W. (1992) Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography.

McRee, D. E. (1999) *J. Struct. Biol.* **125,** 156-165.

Merritt, E. A. & Murphy, M. E. P. (1994) *Acta Cryst.* **D50**, 869–873.

Miyazaki, Y., Matsuo, T. & Suga, H. (2000) *J. Phys. Chem.* **B104,** 8044-8052.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997) *Acta Cryst.* **D53,** 240–255.

Nakasako, M. (1999) *J. Mol. Biol.* **289,** 547-564.

Perrakis, A., Morris, R. M. & Lamzin, V. S. (1999) *Nature. Struct. Biol.* **6**, 458-463.

Pflugrath, J. W. (1999) *Acta Cryst.* **D55,** 1718-1725.

Ravelli, R. B., McSweeney & S. M. (2000) *Structure Fold. Des.* **8,** 315-328.

Sheldrick, G. M. & Schneider, T. R. (1997) *Methods. Enzymol.* **276,** 319-343.

Teng, T. (1990) *J. Appl. Cryst.* **23,** 387-391.

Walsh, M. A., Schneider, T. R., Sieker, L. C., Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1998) *Acta Cryst.* **D54,** 522-546.

Weik, M., Berges, J., Raves, M. L., Gros, P., McSweeney, S., Silman, I., Sussman, J. L., Houee-Levin, C. & Ravelli, R. B. (2002) *J. Synchrotron Radiation* **9,** 342-346.

Weik, M., Ravelli, R. B., Kryger, G., McSweeney, S., Raves, M. L., Harel, M., Gros, P., Silman, I., Kroon, J. & Sussman, J. L. (2000) *Proc. Natl. Acad. Sci.* USA **97,** 623-628.

Chapter 5: Discussion

5.1 Discussion

Methodological improvements in x-ray crystallography along with the world wide scientific shift in focus from genomics to proteomics are the driving forces of the amazing surge in the numbers of structures deposited in the PDB (Berman *et al.*, 2000) in the past few years. The theme of methodological improvements ties all three published papers that comprise this thesis together. All three papers confront important methodological issues. The first two propose new techniques to widen protein crystallography's crystallization bottlenecks, "When less is more: a more efficient vapor diffusion protocol" (Chapter 2, Dunlop and Hazes, 2003) introduces a method of reducing protein consumption. "A modified vapor diffusion crystallization protocol that uses a common dehydrating agent" (Chapter 3, Dunlop and Hazes, 2005), introduces both a novel crystallization plate design that has the potential to solve some of the problems encountered during crystal detection by automatic classification, and a modified vapor diffusion crystallization method that may prove to be more successful than the unmodified variation. The third paper "Pros and cons of cryocrystallography: should we also collect a room-temperature data set?" (Chapter 4, Dunlop *et al.*, 2005) delves into the effect of new methodologies on the structural outcome, in particular it compares the benefits of increased resolution at synchrotron radiation sources with the drawbacks of cryocooling induced structural changes. In this discussion we will look in particular at data and published work that appeared subsequent to publication of the papers presented in this thesis. In addition, attention will be given to future studies that should be considered based on the work presented herein.

5.2 When Less Is More: A More Efficient Vapor Diffusion Protocol

"When less is more: a more efficient vapor diffusion protocol" introduces the 'dilution method', a crystallization method that was designed first and foremost to reduce protein consumption without the need for expensive crystallization robots. This is achieved by diluting the protein and crystallization solutions prior to experimental setup while the reservoir solution remains undiluted, then following evaporation of the excess water from the drop the method becomes equivalent to nanovolume crystallization experiments.

In parallel to our work on the dilution method, Yee (2003) published an alternative approach using a handheld nanoject pipettor (Drummond Co.) to reduce protein consumption without need for a crystallization robot. This technique uses 25-well crystallization slides that fit into standard 24-well plates and the drops are dispensed under oil to prevent evaporation (Yee, 2003). Although this method does achieve comparable protein reduction to the dilution method, and like the dilution method doesn't require the use of crystallization robots, the dilution method is favorable because it is less complex and does not require specialized slides, nanopipettors, or oil. Furthermore, the dilution method can also be used during automated screening in which dispensing accuracy for small drop sizes is increased due to the decreased viscosity of diluted solutions. Also, evaporation of small drops becomes less of an issue.

The dilution method can also be used as an optimization procedure. Because an extra region of crystallization space is explored, there is the potential for crystallization of novel crystal forms, increased crystal size, and a reduction in the number of nucleations (see chapter 2). The advantage of multiple crystal forms is that they each may have quite different diffraction characteristics and different potentials for further optimization. Indeed, based on feedback from both academic and industrial researchers, the dilution method seems to be used most often for optimization (N. Chayen and Vertex Pharmaceuticals Incorporated , personal communication). In our own laboratory the method has been used to optimize conditions for the single-stranded DNA binding protein (ssDNAbp) from vaccinia virus (Brigley and Hazes, unpublished results). The

dilution method produced large well-shaped crystals compared to microcrystalline precipitate without dilution (figure 5.1). Because of its many advantages and beautiful simplicity, the widespread implementation of the dilution method for both manual and automated screening would be a positive evolution. It will be interesting to see what the future holds for since it will only be in its wide scale implementation that its true potential will be revealed.

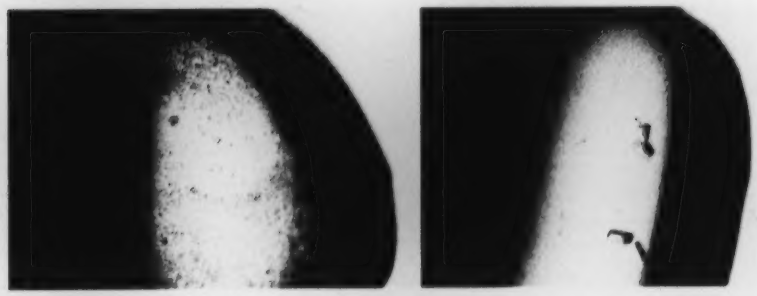

Figure 5.1. The dilution method at work. (a) Microcrystalline precipitate of ssDNAbp produced using the traditional vapor diffusion method. (b) Crystals of ssDNAbp produced using the dilution method.

5.3 A Modified Vapor Diffusion Protocol That Uses A Common Dehydrating Agent

"A modified vapor diffusion protocol that uses a common dehydrating agent" introduces the shared reservoir solution (SRS) method and SRS plate. The SRS method, unlike the traditional vapor diffusion method, uses one reservoir solution as the dehydrant of many different crystallization drops during the crystallization process. The main goal of this work was to design of a new crystallization plate that would give clearer images during automatic imaging and greater crystal detection accuracy during automatic classification.

During automatic imaging and classification the wells and ridges present in the small wells of a 96-well plate distort drop position and shape and thus imaging quality. To attain the goal of giving clearer images and greater crystal detection accuracy, one of our objectives was to create large drop platforms that can contain the drop without the need

for concave wells or ridges. Using the SRS method, wells no longer need to be separated allowing us to combine all wells into a single chamber in our prototype plate. In addition, because the sitting drop platform is totally flat, a number of materials such as glass, kapton tape, ARseal® 90404 (a non-polarizing tape from Adhesives Research Inc.), and plastic may be used in its construction. Compared with plastic, the glass, kapton tape, and ARseal® 90404 have several advantages. Firstly, they do not contain the straight edged flow lines that are often unavoidably introduced during plastics manufacture. These flow lines are confused for the straight edges of crystals during automatic classification. Secondly, glass, kapton tape, and ARseal® 90404 are all non-birefringent materials. This property will allow birefringence to be used as a means of crystal detection when these materials are used for both the sitting drop platform and the lid. Of the three materials, Kapton tape and ARseal® 90404 are the best choice because they are both versatile products ideal for high throughput applications. In addition, kapton tape would be ideal when using the *in situ* X-ray data recording technique, a method of collecting data from crystals while they are still in their crystallization plate (Watanabe *et al.*, 2002), as Kapton is highly transparent to x-rays and displays no fiber diffraction pattern. The large drop platforms also allow drop spacing on the plate to be highly flexible. The common 96-head robots such as the Hydra and Hummingbird must space the drops at regular intervals, however, when using a more flexible crystallization robot such as the Honeybee the drops may be dispensed at any drop spacing, enabling very high densities. The later is important because of the limited high-throughput facility plate storage space. Other practical advantages of the plate include, a less complex set up with only a single reservoir solution required. This is of special advantage when using the RoboDrop combinatorial dispensing approach (Hazes and Price, 2005) where the major problem of dispensing 50 µl amounts of reservoir solution, is alleviated. In addition when using the SRS method the cost of setup is decreased, especially if the expensive commercial screening kits are used, where the requirement for each screening solution becomes only the tiny volume that is used in each crystallization drop.

Steps towards commercial manufacture of the SRS plate are underway. A SRS plate patent, number 60/682,810, has been filed and commercialization is in progress with a

company that specializes in scientific plastics manufacturing. The plates will be made entirely out of plastic and will contain 12 SRS chambers and drop platforms (figure 5.2). This particular commercialization of the SRS plate gives all the benefits of the SRS method and will provide the impetus for its popularization. Other versions of the plate that use glass or tape for the platform and lid may follow later. In addition, specialty plates with one to four chambers and recyclable plates with replaceable drop platforms are a possibility.

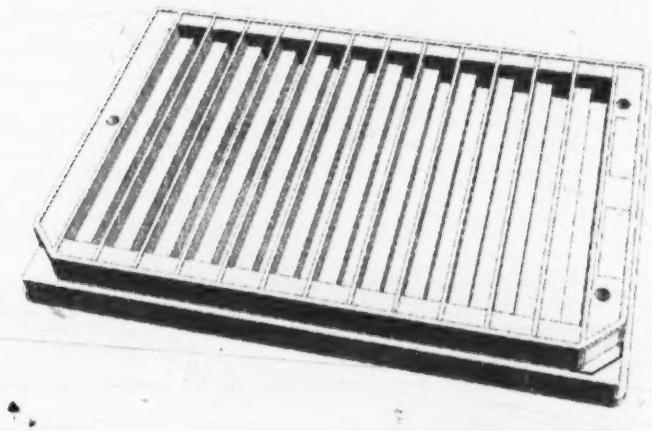

Figure 5.2. SRS plate design (By courtesy of NUNC A/S).

To aid in the wide commercial acceptance of the SRS plates, further tests will be required to determine whether the plates do indeed increase the effectiveness of automatic imaging and classification. A particularly nice experiment would evaluate the crystal detection error rates during automatic classification of drops on either SRS or traditional crystallization plates

The SRS method, from all research performed to date, has been shown to outperform the traditional method in the number of crystallization successes (Dunlop and Hazes, 2005;

Newman, 2005). The reason for this increased success may be related to the dehydrating strength of the reservoir solution, which is related to the measurable parameter called water activity (a_w). Water activity is a measure of the relative availability of water for hydration. It is defined as the ratio of the vapor pressure of water in a material (p) to the vapor pressure of pure water (p_o) at the same temperature.

$$a_w = p / p_o$$

The dehydrating strength of the reservoir solution is an important parameter in all crystallization experiments, however, compared to the traditional method where the water activity of each crystallization drop is determined by its own reservoir solution, all crystallization drops in the SRS method equilibrate to the same water activity, that of the shared reservoir solution. Interestingly, Haas and Drenth have shown that the free energy of binding in protein crystal formation lies in a narrow range (Haas and Drenth, 1995). We hypothesize that the free energy of protein-protein interaction in crystal formation correlates to a large extent with water activity. If this were the case, then the water activity of the reservoir solution would also need to fall within a narrow range, thus explaining the higher success of the SRS method when a precipitant of proper strength is used.

Preliminary investigations to test the water activity of various screening solutions have been undertaken. An AquaLab CX2 Water Activity Meter was used to test the water activity of a number of solutions from the Cryo™ I and Wizard™ I (Emerald BioSystems) screening kits (figure 5.3). We found that water activity decreases nearly linearly as precipitant concentrations increase, as illustrated for NaCl (open circles) in Figure 5.3. Measurements of water activities for several commercial screen conditions show that it is in the range of 0.8-1.0. The great majority have a water activity higher than 0.88 which corresponds to NaCl solutions of 3M or less. Future tests of the hypothesis will include testing the water activities of successful crystallization conditions for proteins, like glucose isomerase, that crystallize in a number of different conditions to determine whether the water activity is similar for all conditions in which they crystallize.

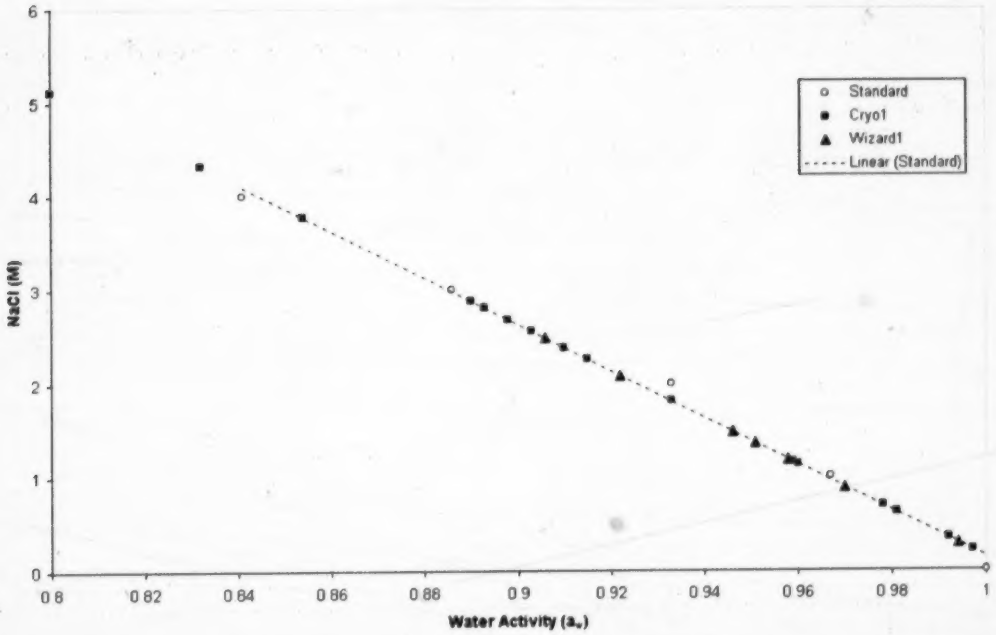

Figure 5.3. Water activity of components of the Cryo I and Wizard I screens (Emerald BioSystems) as plotted on the NaCl water activity standard line.

Although the paper presented in this thesis shows proof of concept of both the SRS plates and SRS crystallization methods, and despite the many advantages of both, a wide acceptance of the plate and method will only come with a large number of positive real life results. More extensive testing of both the plate and method are therefore required. In particular, two features of the SRS method require further investigation to ensure ease of use. The first is the use of volatile reagents, and the second is the dehydrating strength or water activity of the reservoir solution.

5.4 Pros and Cons of Cryocrystallography: Should We Also Collect a Room-Temperature Data Set?

Cryocooling protein crystals prior to data collection is a wide-spread, well respected and successful technique for reducing radiation damage to crystals (see section 1.4.2.3). However, no systematic study of the physiological changes to the structure resulting from

cooling has been undertaken. Such an analysis is complicated because one needs to distinguish direct cryocooling-induced effects from resolution effects, radiation damage effects, and the always present crystal-to-crystal variation. Here we present the first comprehensive study dissecting these effects by comparing 6 PAK pilin structures for which data has been collected from either room or cryo temperature crystals, with different levels of radiation exposure, and to different resolutions ranging between 1.8 and 0.78 Å.

A comparison of the six pilin structures showed that cryocooling-induced structural changes include, firstly, a general shift in atomic coordinates that are largest in the loop regions but can even be detected in the core residues. Secondly, a large increase in detectable bound water molecules with the greatest increase in the second solvation shell. Thirdly, structural variation in the side chain conformations, with long hydrophilic residues such as lysine particularly affected. Since ligand binding and catalysis occur predominantly at the protein surface, it is important to be aware of the potential for cryoinduced effects on sidechain conformations. This resembles the long-standing practice to interpret sidechain conformations and hydration structure in crystal-packing interfaces with extra care.

Cryocooling is often indispensible, as it makes it possible to collect all necessary data using a single crystal thereby avoiding the hazards of merging data sets from many crystals. In addition, cryocooling in combination with brilliant synchrotron x-ray sources results in improvements in resolution and data quality, a factor that also makes it possible to solve protein structures from crystals that were previously considered either to be too small or to diffract too weakly. Nevertheless, cryocooled crystal structures and room temperature structure have their distinct advantages, these advantages being dependent upon the resolutions of the structures. For the pilin structures, the gain in precision for the cryocooled 0.78 Å resolution crystal structure was less than the cryo-induced structural changes, because the 1.63 Å room temperature crystal structure was already excellent. Accordingly, the PAK pilin room temperature structure is more accurate and matches the physiological state of the protein more closely than the cryocooled crystal structure.

However, for a room temperature structure with only moderate resolution, from around 2.5-3 Å, a boost to better than 2 Å from a cryocooled structure could give an increase in precision that is larger than the cryo-induced rms shift of atomic positions.

The ultra-high 0.78 Å resolution PAK pilin structure exposed some interesting issues surrounding the biological relevance of questing for ultra-high resolutions, as no new biologically relevant data was gained as compared to the 1.63 Å structure, despite the significantly improved electron density maps. A literature search reveals that there are instances in which higher resolution structures do yield extra, important biologically relevant information. For example both the serine protease *Bacillus lentus* Subtilisin structure solved to 0.78 Å and the *Tritirachium album limber* proteinase K structure solved to 0.98 Å, resolved short hydrogen bonds in the catalytic site that were not resolved at lower resolutions (Kuhn *et al.*, 1998; Betzel *et al.*, 2001). This shows that ultimately, if the extra detail, such as polar hydrogens, anisotropic atom motions, and low occupancy atoms, and the extra precision in coordinate distances are relevant for the question being studied, then questing for ultra-high resolution structures is worth it. However, in other cases, like pilin, where there are no questions that rely on extremely detailed features the push for ultra-high resolution is not as pressing, the one caveat being that you can never be certain what the extra detail is going to show until you have tried.

Apart from the need for cryocooling, pushing the resolution limit also requires longer exposure times, thus increasing the likelihood of radiation damage. The specific radiation damage effects on disulphide bonds and other atoms in proteins have been well-documented in a series of papers from 2000 and thereafter (Teng and Moffat, 2000; Ravelli and McSweeney, 2000; O'Neill *et al.*, 2002; Garman, 2003). However, the pilin data was collected prior to this and as a result the data collection strategy was not optimal. Instead of collecting the weak signal high resolution data first, a process that takes a long time and exposes the crystal to a high radiation dose, the strongest reflection low resolution data should have been collected first. This is because the data collection time for low resolution reflections is relatively short thus exposing the crystal to little radiation damage, yet the reflections because of their high intensity are significant

contributers to the final structure. Additionally, instead of increasing dose by lengthening exposure time per image it might have been better to increase the number of images. Data quality would then be improved by increased redundancy and data that was significantly radiation damaged could be deleted without affecting data completeness.

The non-specific damage may affect overall electron density quality. In the PAK pilin structure non-specific damage manifested, firstly, as loss of weak signals such as H-atoms. The 0.78 Å resolution structure showed experimental electron density for many fewer H atoms compared to other similarly high resolution structures, and in comparison to the first least damaged data batch. Secondly, the non-specific damage manifested as an increase in R_{merge}. It is obvious, therefore, that optimal electron density quality requires a balance between maximizing resolution and minimizing radiation damage.

Given the extensive radiation damage at modern third generation synchrotrons, the question of how to design data collection strategies to minimize radiation damage while still maximizing resolution arises. Some suggestions are, firstly, that an estimate of the maximum dose should be made so as to allow data collection to be planned in a way that does not exceed this while still maximizing the information obtained. However, this calculation does not hold for every crystal as the actual lifetime of the crystal could become shorter in cases where specific damage breaks down crystal contacts (Murray *et al.*, 2005). Secondly, the use of radical scavengers such as ascorbate may minimize radiation damage (Murray and Garman, 2002). Although no recent data has become available, in at least one instance the data collected from a lysozyme crystal soaked in an ascorbate-containing cryosolution showed a smaller differential relative increase in B-factors over successive datasets and improved electron-density difference maps. Thirdly, if crystals are larger than the beam they should be translated to spread the radiation dose over multiple spots on the crystal.

If you still desire to push the diffraction resolution of your crystal to its limit there is a new and very promising method that offers the possibility of correcting the data for radiation damage. This method, named zero-dose extrapolation, can computationally

correct individual intensities that have been altered by radiation damage back to their zero dose level using multiple measurements of each reflection (Diederichs *et al.*, 2003). Methods are even being developed to use the specific radiation damage to disulphide bonds to obtain phase information. In the radiation damage-induced phasing method (RIP) a low dose data set is collected followed by a high dose 'burn' data set, and a second low dose data set. The intensity differences between the two low dose data sets are used to determine the sulphur atom substructure responsible for the isomorphous difference signal (Banumathi *et al.*, 2004).

5.5 Conclusion

Protein crystallography has undergone a vigorous transformation in the new millennium with the realization that detailed structure-based analysis of gene products is the logical progression following the completion of genome sequencing projects. X-ray crystallography is now being used by an ever increasing population of researchers to solve structures at an ever increasing pace. Not only are the number of structures deposited in the PDB increasing exponentially, but increasingly complex problems are being addressed, with the major success being the structure of the ribosome in 2000 (Ban *et al.*, 2000; Wimberley *et al.*, 2000; Schleunzen *et al.*, 2000). Methodological improvements have been instrumental in achieving this progress. From highly sophisticated advances such as third generation synchrotrons, and high-throughput robots, to simple but critical innovations such as the cryoloop, all contribute to the improved ease and increased throughput of protein structure solution. All the papers presented in this thesis have addressed important methodological issues facing protein crystallography today, from reducing protein consumption during the crystallization process, to increasing the success of both crystallization and crystal detection, to delving into the effect of crystal cryocooling on the final structure. It is hoped that the methodological studies presented here will contribute to the continued success of protein crystallography.

5.6 References

Ban, N., Nissen, P., Hansen, J., Moore, P.B., and Steitz, T.A. (2000). *Science*. **289**, 905–920.

Banumathi, S., Zwart, P.H., Ramagopal, U.A., Dauter, M., and Dauter, Z. (2004). *Acta Cryst. D***60**, 1085-93.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. 2000. *Nucleic Acids Res.* **28**, 235-242.

Betzel, C., Gourinath, S., Kumar, P., Kaur, P., Perbandt, M., Eschenburg, S., and Singh, T.P. 2001. *Biochemistry*. **40**, 3080-3088.

Diederichs, K., McSweeney, S., and Ravelli, R.B.G. (2003). *Acta Cryst.* **D59**, 903-909.

Dunlop, K.V., and Hazes, B. (2003). *Acta Cryst.* **D59**, 1797-1800.

Dunlop, K.V., and Hazes, B. (2005). *Acta Cryst.* **D61**,1041-1048.

Dunlop, K.V., Irvin, R.T., and Hazes, B. (2005). *Acta Cryst.* **D61**, 80-87.

Garman, E. (2003). *Curr Opin Struct Biol.* **13**, 545-551.

Haas, C., and Drenth, J. (1995). *J Cryst Growth.* **154,** 126-135.

Kuhn, P., Knapp, M., Soltis, S.M., Ganshaw, G., Thoene, M., and Bott, R. 1998. *Biochemistry*. **37**, 13446-13452.

Murray, J., and Garman, E. 2002. *J Synchrotron Radiat.* **9**, 347-54.

Murray, J.W., Rudino-Pinera, E., Owen, R.L., Grininger, M., Ravelli, R.B., and Garman, E.F. 2005. *J Synchrotron Radiat.* **12**, 268-275.

Newman, J. (2005). *Acta Cryst.* D**61**, 490-493.

O'Neill, P., Stevens, D.L., and Garman, E.F. (2002). *J. Synchrotron Rad.* **9**, 329-332.

Ravelli, R.B. and McSweeney, S.M. (2000) *Struct. Fold. Des.* **8**, 315-328.

Schluenzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., Franceschi, F., and Yonath, A. (2000). *Cell.* **102**, 615–623.

Teng, T., and Moffat, M. (2000) *J. Synchrotron Rad.* **7**, 313-317.

Watanabe, N., Murai, H., and Tanaka, I. (2002). *Acta Cryst.* D**58**, 1527-1530.

Wimberly, B.T., Brodersen, D.E., Clemons, W.M. Jr., Morgan-Warren, R.J., Carter, A.P., Vonrhein, C., Hartsch, T., and Ramakrishnan, V. (2000). *Nature.* **407**, 327–339.

Yeh, J.I. (2003). *Acta Cryst.* D**59**, 1408-1413.

Appendix

2Fo –Fc electron density for cysteine residues 129 and 142 contoured at 1.5 σ.

(A) Density for structure 2, based on rotating anode diffraction data to 1.9 Å resolution from a second generation source, shows two conformations of the disulphide bond. (B) Density for the first 500 images of the structure 1 data set, based on 0.78 Å diffraction data from a third-generation synchrotron source, shows two conformations for the sulfur atom of cysteine 142. The density for the second conformation of the sulfur atom is much less for both A and B in comparison to corresponding density for the complete data set 1. Carbon, nitrogen, and sulfur atoms are displayed in yellow, blue, and green, respectively. Electron density is purple, positive and negative electron density are green and red respectively.